Nº 98-306-X2016001 au catalogue
ISBN 978-0-660-26770-8

Rapport technique du recensement de la population

Rapport technique sur l'échantillonnage et la pondération

Recensement de la population, 2016

Date de diffusion : le 11 septembre 2018

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-514-283-9350

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « Normes de service à la clientèle ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

.	indisponible pour toute période de référence
..	indisponible pour une période de référence précise
...	n'ayant pas lieu de figurer
0	zéro absolu ou valeur arrondie à zéro
0^s	valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
^{p}	provisoire
^{r}	révisé
x	confidentiel en vertu des dispositions de la *Loi sur la statistique*
^{E}	à utiliser avec prudence
F	trop peu fiable pour être publié
*	valeur significativement différente de l'estimation pour la catégorie de référence (p<0,05)

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté la Reine du chef du Canada, représentée par le ministre de l'Industrie 2018

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'entente de licence ouverte de Statistique Canada.

Une version HTML **est aussi disponible.**

This publication is also available in English.

Table des matières

Introduction 1

1. Collecte des données du recensement 2
   - 1.1 Méthodologie de collecte 2
   - 1.2 Questionnaires et formules 6
   - 1.3 Taux de réponse à la collecte 6

2. Échantillonnage 7
   - 2.1 Univers pour l'échantillon du questionnaire détaillé 7
   - 2.2 Plan d'échantillonnage pour le questionnaire détaillé 7

3. Traitement des données du recensement 8
   - 3.1 Introduction 8
   - 3.2 Réception et enregistrement 8
   - 3.3 Imagerie et saisie à partir d'images 8
   - 3.4 Contrôles de couverture, contrôles d'intégralité et suivi des questionnaires rejetés au contrôle 9
   - 3.5 Codage 10
   - 3.6 Classification et ajustements pour la non-réponse des logements inoccupés et des logements non répondants 11
   - 3.7 Obtention des données sur le revenu 11
   - 3.8 Non-réponse 12
   - 3.9 Vérification et imputation 12
   - 3.10 Pondération 13
   - 3.11 Taux de réponse finaux 13

4. Estimation à l'aide de l'échantillon du questionnaire détaillé du recensement 15
   - 4.1 Éléments à considérer pour choisir une méthode d'estimation 15
   - 4.2 Régions de pondération 16
   - 4.3 Poids de sondage 22
   - 4.4 Ajustement pour la couverture et pour la non-réponse totale 22
   - 4.5 Calage final 24
   - 4.6 Détails sur la sélection des contraintes 26

5. Évaluation des procédures de pondération 32
   - 5.1 Répartition des poids 32
   - 5.2 Écarts entre les chiffres du recensement et les estimations pour le questionnaire détaillé, Canada 34

6. Estimation de la variance 45
6.1 Éléments à considérer pour choisir une méthode d'estimation de la variance 46
6.2 Estimateur de variance 47
6.3 Ajustement des poids de répliques 48

7. Conclusion 50

Annexe A – Glossaire 51

Annexe B – Historique de l'échantillonnage dans le cadre du recensement canadien 54

Annexe C – Contraintes utilisées ou exclues dans le processus de pondération 56

Annexe D – Wood Buffalo 76

Bibliographie 85

Rapport technique sur l'échantillonnage et la pondération

Introduction

Dans le cadre du Programme du Recensement de 2016, les ménages canadiens sont dénombrés à l'aide de deux principaux types de questionnaires : le questionnaire abrégé et le questionnaire détaillé. En 2016, un échantillon de 25 % des ménages canadiens a reçu un questionnaire détaillé, qui comprenait aussi les questions du questionnaire abrégé. Les autres ménages ont reçu un questionnaire abrégé. En 2011, le gouvernement avait décidé d'abandonner le caractère obligatoire du questionnaire détaillé du recensement. Il l'a rétabli pour le Recensement de 2016.

Outre les questions du questionnaire abrégé, le questionnaire détaillé comprend un ensemble de questions visant à dresser un portrait complet de la population et des ménages canadiens, selon leurs caractéristiques démographiques, sociales et économiques.

Les estimations produites à partir de réponses à des questions communes aux deux types de questionnaires sont obtenues de l'ensemble de la population par un recensement. En effet, tous les ménages répondants aux deux types de questionnaires contribuent alors à un chiffre donné. C'est le cas, par exemple, du chiffre de population pour un groupe d'âge donné.

Les estimations produites à partir de réponses à au moins une question propre au questionnaire détaillé sont obtenues de l'échantillon. Dans ce cas, seuls les ménages répondants de l'échantillon du questionnaire détaillé contribuent à l'estimation. L'estimation du taux de chômage ou celle de la population selon le plus haut niveau de scolarité atteint sont des exemples de telles estimations.

L'échantillon du questionnaire détaillé est réparti uniformément au niveau géographique afin d'assurer une bonne fiabilité des estimations pour toutes les régions du pays et d'accorder la même importance relative à toutes les unités géographiques d'une taille donnée.

Le présent rapport technique porte sur la méthodologie utilisée dans le processus d'estimation à partir de l'échantillon du questionnaire détaillé du Recensement de la population de 2016. Le chapitre 1 décrit en détail les méthodes de collecte utilisées dans le cadre du recensement et de l'échantillon du questionnaire détaillé. Le chapitre 2 décrit comment l'échantillonnage est mis en œuvre pour le questionnaire détaillé. Le chapitre 3 explique les procédures de traitement des données. Le chapitre 4 donne un aperçu des procédures utilisées pour assigner un poids aux unités répondantes de l'échantillon du questionnaire détaillé afin d'obtenir des estimations pour la population. Le chapitre 5 porte sur diverses évaluations des procédures de pondération, tandis que le chapitre 6 donne un aperçu de la méthodologie d'estimation de la variance dans le cadre de l'échantillon du questionnaire détaillé du Recensement de 2016. Une conclusion est présentée au chapitre 7.

1. Collecte des données du recensement

L'étape de la collecte des données du Recensement de 2016 visait à faire en sorte que chacun des 15,4 millions de logements du Canada soit dénombré et que, pour chaque logement occupé, le ménage correspondant remplisse un questionnaire du recensement. Le recensement visait l'ensemble de la population du Canada, qui comprend les citoyens canadiens (de naissance ou par naturalisation), les immigrants reçus et les résidents non permanents, ainsi que les membres de leur famille vivant sous le même toit. Les résidents non permanents sont les personnes qui vivent au Canada et qui sont titulaires d'un permis de travail ou d'études, ou qui demandent le statut de réfugié, ainsi que les membres de leur famille qui vivent avec eux.

Le recensement dénombrait également les citoyens canadiens et les immigrants reçus qui étaient temporairement à l'extérieur du pays le jour du recensement, y compris les fonctionnaires fédéraux et provinciaux ou territoriaux en poste à l'extérieur du Canada, le personnel des ambassades canadiennes à l'étranger, les membres des Forces canadiennes en poste à l'étranger et les Canadiens qui se trouvaient à bord de navires marchands.

1.1 Méthodologie de collecte

Depuis le Recensement de 2011, une nouvelle méthodologie de collecte est utilisée pour la collecte du recensement. Appelée méthodologie par vagues, cette méthode consiste à communiquer avec les ménages non répondants à des moments clés afin de leur rappeler de participer au recensement et de les inciter à remplir le questionnaire. À chaque vague de contact, les renseignements nécessaires à la réponse sont fournis aux ménages. Cette méthode, qui s'inscrit dans un contexte où chaque ménage canadien est tenu par la loi de répondre aux questions du recensement, est conçue de façon à encourager la réponse hâtive par Internet tout en atténuant le risque de diminution de la réponse totale et le besoin de suivis coûteux en personne.

Cette méthodologie diffère selon le mode de collecte utilisé pour distribuer les documents du recensement pour une région donnée[1]. Ces modes de collecte sont décrits à la section suivante. En 2016, les ménages canadiens pouvaient répondre par Internet, en remplissant un questionnaire papier (retour par la poste) ou en communiquant avec l'Assistance téléphonique du recensement.

1.1.1 Modes de collecte : livraison des questionnaires du recensement

Les trois modes de collecte utilisés au recensement canadien sont l'envoi par la poste, le listage/livraison et le recensement par interview. Afin d'organiser la collecte du recensement le plus efficacement possible, le territoire du Canada est divisé en petites unités géographiques appelées unités de collecte (UC). Au Recensement de 2016, on comptait environ 46 000 UC au Canada. Il y a cinq types d'UC et un des trois modes de collecte est assigné à chacun. Ainsi, le même mode de collecte est utilisé pour tous les logements d'une UC.

1. La collecte des données pour la subdivision de recensement de Wood Buffalo fait exception à cette méthodologie. Le 1er mai 2016, un incendie de forêt s'est déclaré dans le sud-ouest de Fort McMurray, en Alberta. Deux jours plus tard, soit le 3 mai, il a ravagé la région, détruisant bon nombre de maisons et d'immeubles. Statistique Canada a mis en place une stratégie alternative dans la subdivision de recensement de Wood Buffalo, qui englobe Fort McMurray, en raison de l'évacuation des résidents durant la période de collecte. Plus de détails à propos de la méthode de collecte mise en place dans la subdivision de recensement de Wood Buffalo sont disponibles à l'annexe D.

Tableau 1.1.1.1
Mode de collecte selon le type d'unité de collecte

Type d'UC	Mode de collecte
Envoi par la poste	Envoi par la poste
Listage/livraison	Listage/livraison
Recensement par interview	Recensement par interview
Logements collectifs	Listage/livraison
Réserves	Recensement par interview

UC = Unité de collecte.

Les différents modes de collecte sont décrits ici.

1.1.1.1 Envoi par la poste

Dans les UC dont le mode de collecte est l'envoi par la poste, on a recours au système postal pour envoyer le matériel du recensement. Ce type d'envoi permet une distribution efficace et coordonnée, sans avoir à recruter et à former un effectif important d'agents recenseurs. Les UC d'envoi par la poste sont typiquement dans les régions urbaines. Bien que ce type d'UC couvre maintenant environ 82 % des logements au Canada, il représente seulement une très petite partie de la superficie du territoire. Le quart des logements privés provenant de ces UC sont sélectionnés pour recevoir un questionnaire détaillé; les autres logements reçoivent un questionnaire abrégé.

1.1.1.2 Listage/livraison

Les UC dont le mode de collecte est le listage/livraison sont typiquement des régions rurales. Dans ces régions, des agents recenseurs effectuent le listage des logements et font la livraison du matériel du recensement. Environ 17 % des logements du Canada se trouvent dans ce type d'UC, qui couvre une partie importante de la superficie du territoire. Le quart des logements privés provenant de ces UC sont sélectionnés pour recevoir un questionnaire détaillé. Ce mode de collecte est aussi utilisé pour les UC de logements collectifs.

1.1.1.3 Recensement par interview

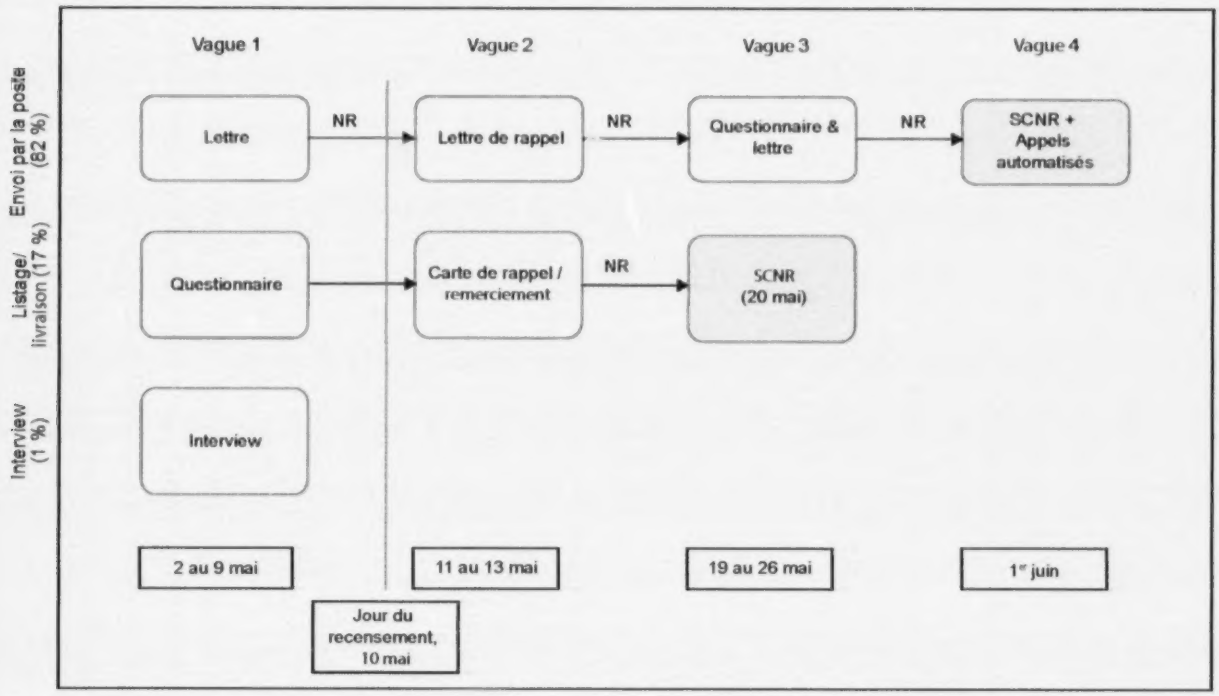
Les UC dont le mode de collecte est le recensement par interview sont soit des réserves indiennes, soit des UC situées dans des endroits éloignés ou difficiles d'accès. Afin de limiter le nombre de déplacements souvent coûteux et logistiquement complexes des agents recenseurs pour faire des suivis à ces endroits, ces derniers ne se limitent pas à faire le listage des logements, mais remplissent aussi immédiatement le questionnaire avec chacun des ménages. Ce type d'UC couvre un peu plus de la moitié de la superficie du territoire du Canada, mais seulement environ 1 % des logements. Tous les logements privés de ces UC reçoivent un questionnaire détaillé.

1.1.2 Méthodologie par vagues du recensement

La méthodologie par vagues a été conçue de façon à encourager la réponse par Internet tout en offrant une alternative aux ménages qui ne souhaitent pas remplir leur questionnaire par Internet. Cette approche par vagues comporte de nombreux avantages pour minimiser la non-réponse et augmenter le nombre de réponses par Internet, ce qui accélère l'enregistrement des questionnaires, facilite l'enchaînement des questions et la saisie des données, et ultimement augmente la qualité.

La méthodologie par vagues s'applique de façon différente selon le mode de collecte. Les UC ont donc été regroupées selon le mode de collecte, et une variante de la méthodologie par vagues a été développée pour chacun des trois groupes. En raison de la nature même des UC de recensement par interview, une méthodologie à une seule vague a été adoptée pour ces unités. Les sections suivantes ainsi que la figure 1.1.2.1 donnent un aperçu de la méthodologie par vagues qui a été utilisée pour le Recensement de 2016.

Figure 1.1.2.1
Aperçu de la méthodologie par vagues du Recensement de 2016



NR = non-répondants
SCNR = suivi des cas de non-réponse
Source : Statistique Canada, Recensement de la population, 2016.

1.1.2.1 Unités de collecte d'envoi par la poste

Dans les UC d'envoi par la poste, la première vague a consisté en l'envoi d'une lettre seulement et dans laquelle on demandait aux ménages de remplir le questionnaire en ligne au moyen du code d'accès sécurisé (CAS) qui leur était fourni ou d'utiliser un système automatisé accessible à partir d'un numéro de téléphone sans frais pour demander qu'un questionnaire papier leur soit envoyé par la poste. Les lettres de la vague 1 ont été livrées par le système postal une semaine avant le jour du recensement, soit le 2 mai 2016.

La deuxième vague a consisté en une lettre de rappel envoyée à tous les ménages non répondants de la vague 1 à une certaine date selon la région. Cette lettre rappelait aux ménages qu'ils étaient tenus par la loi de répondre au recensement. Elle comprenait aussi le CAS et le numéro de téléphone sans frais, tout comme la lettre de la première vague. Elle a été envoyée aux ménages du 11 au 13 mai.

À la troisième vague, du 19 au 26 mai, on a envoyé un questionnaire papier aux ménages non répondants. Les ménages pouvaient toujours répondre en ligne au moyen d'un CAS fourni sur la page couverture du questionnaire. De plus, une lettre accompagnait le questionnaire et mentionnait que si le questionnaire n'était pas dûment rempli au 31 mai 2016, un agent recenseur allait communiquer avec le ménage par appel téléphonique ou en personne pour remplir le questionnaire. La lettre mentionnait aussi que si le ménage refusait de répondre aux questions du recensement, le dossier pouvait être déféré au Service des poursuites pénales du Canada, qui prendrait alors les mesures qui s'imposent en vertu de la *Loi sur la statistique*.

La quatrième vague, qui a débuté le 1er juin 2016, a été le suivi des cas de non-réponse (SCNR) sur le terrain, et incluait aussi un rappel téléphonique automatisé au début de la période de suivi. Le SCNR est décrit à la section 1.1.3 de ce document.

1.1.2.2 Unités de collecte de listage/livraison

Dans les UC des secteurs de listage/livraison, une variante de la méthodologie par vagues a été utilisée. À la première vague, des agents recenseurs ont livré un questionnaire papier à tous les logements dans ces UC entre le 2 et le 9 mai. Ce questionnaire était assorti également d'un CAS, donnant ainsi l'option au ménage de répondre par Internet.

À la deuxième vague, du 11 au 13 mai, tous les logements dans ces UC ont reçu par la poste une carte de remerciement/rappel, peu importe qu'ils aient déjà répondu ou non, car il n'est généralement pas possible dans ces régions de faire un envoi postal à des adresses de voirie précises sans le nom de l'occupant.

Finalement, la dernière vague consistait à passer directement au suivi des cas de non-réponse sur le terrain à partir du 20 mai.

1.1.3 Suivi des cas de non-réponse

Comme mentionné à la section précédente, la dernière vague de la méthodologie par vagues correspond au SCNR. À cette étape, les agents recenseurs appellent et visitent les ménages qui n'ont pas répondu au questionnaire. Chaque ménage non répondant pour lequel un numéro de téléphone était disponible a reçu un rappel téléphonique automatisé au début de la période du SCNR. Ce message rappelait aux non-répondants leur obligation légale de répondre au questionnaire du recensement.

Le SCNR a été mené en personne par des agents recenseurs ou au téléphone par des agents de centres téléphoniques. Les employés du recensement avaient à leur disposition des informations pour gérer leur travail dans le Portail de gestion de la collecte. Ce portail, accessible par Internet, est un système informatique qui a aussi facilité la cueillette de renseignements sur les progrès et les coûts de la collecte.

1.1.4 Vérification de l'occupation des logements

Avant le SCNR, des opérations sur le terrain ont également eu lieu afin d'effectuer la vérification de l'occupation des logements (VOL) non répondants. Cette activité qui commençait peu de temps après le jour du recensement, le 10 mai 2016, avait pour but de trouver le plus grand nombre possible de logements inoccupés au jour du recensement ou annulés (adresses qui ne sont pas des logements privés ou collectifs) avant le début du SCNR. En identifiant ces logements à une date proche du jour du recensement, la classification du statut d'occupation devrait être plus précise et même plus facile à effectuer. De plus, la VOL permet de réduire la charge de travail du SCNR puisque les logements inoccupés ou annulés identifiés par cette vérification n'ont pas besoin de suivi.

Tout de même, des erreurs de classification d'un logement comme étant occupé ou inoccupé se produisent durant la VOL et le SCNR. Certains logements classés comme étant inoccupés sont en fait occupés et certains logements non répondants sont des logements inoccupés. Par conséquent, une autre opération, l'Enquête sur la classification des logements (ECL), se déroule après le SCNR. Elle détermine le statut d'occupation d'un échantillon de logements pour lesquels aucun questionnaire rempli n'a été reçu (logements inoccupés, logements non répondants ou cas non résolus) et elle estime les taux d'occupation de l'ensemble des logements du pays dans la même situation. À partir des résultats de l'enquête, on ajuste les chiffres du Recensement de la population durant le traitement. Plus de détails se rattachant à l'ECL sont décrits à la section 3.6.

1.2 Questionnaires et formules

Au Recensement du Canada, divers questionnaires et formules sont utilisés pour recueillir les données, et les répondants ont l'option de répondre au questionnaire du recensement par Internet ou en remplissant un questionnaire papier.

Le Registre des visites (RV) sert à dresser la liste des logements privés et collectifs occupés et inoccupés ainsi que des exploitants et exploitations agricoles qui se trouvent dans des régions du pays où l'on effectue le listage des logements sur place avant de livrer les questionnaires (c.-à-d. dans les unités de collecte des secteurs de listage/livraison; voir la section 1.1.1.2). Le RV sert de liste d'adresses aux fins du contrôle et des opérations sur le terrain dans le cadre de la collecte des données.

Le questionnaire 2A (http://www23.statcan.gc.ca/imdb/p3Instr_f.pl?Function=getInstrumentList&Item_Id=295241&UL=1V&) est le questionnaire abrégé de base servant à dénombrer les logements privés. Chaque ménage est invité à y dénombrer tous ses membres qui sont inclus dans la population du recensement et à répondre aux questions pour eux.

Les questionnaires 2A-L (http://www23.statcan.gc.ca/imdb/p3Instr_f.pl?Function=getInstrumentList&Item_Id=295122&UL=1V&) et 2A-R (http://www23.statcan.gc.ca/imdb/p3Instr_f.pl?Function=getInstrumentList&Item_Id=295299&UL=1V&) correspondent au questionnaire détaillé. Les questions du questionnaire 2A sont également sur ces formules, suivies de questions supplémentaires sur des sujets variés tels que la scolarité, l'emploi et le logement. Le questionnaire 2A-L est administré à 25 % des ménages canadiens demeurant dans les secteurs d'envoi par la poste et les secteurs de listage/livraison tels que décrits aux sections 1.1.1.1 et 1.1.1.2, soit les ménages de l'échantillon du questionnaire détaillé. Le questionnaire 2A-R contient presque toutes les mêmes questions que le questionnaire 2A-L et est utilisé dans les régions dont le mode de collecte est le recensement par interview. Tous les ménages de ces régions font partie de l'échantillon du questionnaire détaillé.

Le questionnaire 2C (http://www23.statcan.gc.ca/imdb/p3Instr_f.pl?Function=getInstrumentList&Item_Id=295305&UL=1V&) sert à dénombrer les Canadiens en poste à l'étranger. Le questionnaire 2C contient les mêmes questions que la formule 2A.

Le questionnaire 3A (http://www23.statcan.gc.ca/imdb/p3Instr_f.pl?Function=getInstrumentList&Item_Id=295255&UL=1V&) est un questionnaire individuel du recensement qui sert à dénombrer les personnes dans les logements collectifs. Il peut également être utilisé pour dénombrer des résidents habituels d'un ménage privé qui veulent remplir leur propre questionnaire du recensement plutôt qu'être inscrits sur un questionnaire 2A.

1.3 Taux de réponse à la collecte

Le taux global de réponse à la collecte du Recensement de la population de 2016 est de 98,4 %. Ce taux est produit directement des résultats de la collecte (en date du 4 août 2016), c'est-à-dire avant que le traitement et la vérification de la qualité des données ne soient terminés. Il représente le nombre de logements privés pour lesquels un questionnaire a été retourné, divisé par le nombre de logements privés codés comme occupés par les agents recenseurs. Le taux de réponse à la collecte au sein de l'échantillon du questionnaire détaillé est quant à lui de 97,8 % (pour plus de détails, consultez le taux de réponse de la collecte du Recensement de la population de 2016 (http://www12.statcan.gc.ca/census-recensement/2016/ref/Taux-reponse-fra.cfm)).

2. Échantillonnage[2]

Lorsqu'une enquête-échantillon est menée, il est important de bien planifier la sélection de l'échantillon. L'échantillonnage est le processus qui permet de choisir un sous-ensemble de la population cible de l'enquête qui recevra le questionnaire et dont les réponses permettront de produire des inférences sur l'ensemble de la population. On distingue deux types d'échantillonnage : l'échantillonnage probabiliste et l'échantillonnage non probabiliste. Lorsqu'il est important de produire des inférences statistiques sur la population entière, un échantillon probabiliste est préférable, car il est alors possible de calculer la probabilité de sélection des unités et d'estimer l'erreur d'échantillonnage. Le présent chapitre porte sur la sélection de l'échantillon qui a reçu le questionnaire détaillé du Recensement de 2016.

2.1 Univers pour l'échantillon du questionnaire détaillé

L'univers des ménages du recensement a été divisé en trois parties : les ménages privés, les ménages collectifs et les ménages situés à l'extérieur du Canada. L'univers pour l'échantillon du questionnaire détaillé est uniquement formé des ménages privés, y compris les ménages vivant dans des logements privés annexés aux logements collectifs du Canada. Cet univers exclut les réserves indiennes et établissements indiens partiellement dénombrés. Sauf indication contraire, l'expression « dans le champ de l'enquête » sera utilisée pour indiquer qu'un ménage fait partie de l'univers pour l'échantillon du questionnaire détaillé (c.-à-d. les ménages privés qui ne sont pas dans des réserves indiennes et établissements indiens partiellement dénombrés), tandis que l'expression « hors du champ de l'enquête » fait référence aux ménages non compris dans l'univers (c.-à-d. les ménages vivant dans des logements collectifs, ceux à l'extérieur du Canada et ceux vivant dans des réserves indiennes et établissements indiens partiellement dénombrés).

2.2 Plan d'échantillonnage pour le questionnaire détaillé

Dans la vaste majorité des cas, le questionnaire détaillé est envoyé au quart des ménages de l'univers du questionnaire détaillé afin de recueillir des données démographiques et socioéconomiques à propos de la population canadienne. L'échantillon a été tiré de la liste des logements du Recensement de la population de 2016. Au moment de la sélection de l'échantillon, les adresses correspondant à des logements hors du champ de l'enquête étaient inconnues, ce qui signifie que certains logements ont reçu un questionnaire détaillé par erreur. Après avoir établi qu'un logement était hors du champ de l'enquête, toute autre activité de collecte ou de traitement des données a été abandonnée.

Le plan d'échantillonnage utilisé pour la sélection des logements qui reçoivent le questionnaire détaillé est un plan stratifié et systématique. Les strates du plan d'échantillonnage sont les UC. Dans les UC où le mode de collecte est l'envoi par la poste, l'échantillonnage est systématique et la fraction de sondage est de un quart. Le point de départ de la sélection est aléatoire. Dans les UC où le mode de collecte est le listage/livraison, l'échantillonnage est systématique et les logements échantillonnés sont le 4^{e} logement listé, le 8^{e}, le 12^{e}, etc. Par exemple, si l'une de ces UC comporte sept logements, un seul logement est sélectionné. Enfin, dans les UC où le mode de collecte est le recensement par interview, tous les ménages sont sélectionnés. Ces UC sont des strates à tirage complet.

Il y a une exception au plan d'échantillonnage. En effet, les logements privés annexés à des logements collectifs sont ajoutés à l'échantillon avec certitude. Cependant, ils ne remplissent que le questionnaire abrégé. Les réponses au questionnaire détaillé sont par la suite imputées pour ces ménages.

À l'exception des ménages privés annexés à des logements collectifs, on demande aux ménages sélectionnés dans l'échantillon de remplir le questionnaire détaillé du recensement. On demande aux ménages des logements privés qui ne font pas partie de l'échantillon du questionnaire détaillé de remplir le questionnaire abrégé.

2. Pour l'historique de l'échantillonnage dans les recensements du Canada, consulter l'annexe B.

3. Traitement des données du recensement

3.1 Introduction

Le présent chapitre est consacré au traitement de tous les questionnaires remplis, quel que soit leur type, depuis la réception des questionnaires jusqu'à la création d'une base de données du recensement exacte et complète. On trouvera ci-après la description des étapes de l'enregistrement des questionnaires, de l'imagerie et de la saisie des données des questionnaires, du contrôle, de la correction des erreurs, du suivi des questionnaires rejetés au contrôle, du codage, de la classification des logements et des ajustements pour la non-réponse, du couplage des données sur le revenu, de l'imputation, de la pondération et des taux de réponse finaux.

Les processus automatisés, mis en œuvre pour le Recensement de 2016, ont dû être surveillés de manière à ce que toutes les résidences canadiennes soient dénombrées une seule fois. Le Système de contrôle principal (SCP) a été mis au point afin de contrôler et de surveiller l'enchaînement des opérations, de la collecte au traitement des données. Le SCP contenait une liste principale de tous les logements existants au Canada, sur laquelle figurait pour chacun d'eux un identificateur unique. Ce système était mis à jour régulièrement, en fonction des renseignements sur la situation de chaque logement dans le processus d'enchaînement des opérations du recensement (p. ex. livraison, réception ou traitement). Des rapports étaient produits quotidiennement et mis en ligne à la disposition des gestionnaires, afin que les opérations du recensement se déroulent de façon efficace et efficiente.

3.2 Réception et enregistrement

Les réponses reçues par Internet ou recueillies par interview téléphonique par l'Assistance téléphonique du recensement arrivaient directement au Centre des opérations des données (COD) et leur réception était enregistrée automatiquement.

Les répondants qui avaient rempli un questionnaire papier devaient le retourner au COD par la poste. Postes Canada enregistrait automatiquement leur réception dans plusieurs emplacements au Canada (dans le cadre du processus ordinaire de traitement du courrier) en numérisant par la fenêtre transparente des enveloppes-réponses le code à barres qui figurait à la première page du questionnaire. Les enveloppes étaient ensuite livrées au COD chaque jour ouvrable. En outre, Postes Canada envoyait quotidiennement un fichier énumérant tous les questionnaires du recensement reçus à chaque établissement régional de traitement du courrier, selon la date de réception.

L'enregistrement de chaque questionnaire retourné était signalé dans le SCP à Statistique Canada. Le SCP produisait quotidiennement une liste de tous les logements pour lesquels un questionnaire n'avait pas été reçu, et la transmettait aux opérations sur le terrain pour éviter d'effectuer un suivi des cas de non-réponse auprès des ménages qui avaient déjà rempli leur questionnaire.

3.3 Imagerie et saisie à partir d'images

En 2016, tous les questionnaires papier du recensement (2A, 2C, 2A-L, 2A-R, 3A) ont été convertis en images. Le processus d'imagerie comprenait les étapes suivantes :

- Préparation des documents : Les questionnaires retournés par la poste ont été retirés des enveloppes, puis débarrassés de tout corps étranger (comme des trombones ou des agrafes) en vue de la numérisation. Les questionnaires ont été mis en lots selon le type de formule, et divisés en feuilles individuelles en supprimant le dos de la reliure.
- Numérisation : L'opération consistait à convertir les questionnaires en images numériques.
- Évaluation automatisée de la qualité des images : Un système automatisé a analysé les images afin de déceler les erreurs ou les anomalies. Les images rejetées à cette étape ont été transmises pour examen à un opérateur spécialisé en analyse de documents.

- Analyse des documents : À cette étape, les images contenant des anomalies ont été présentées à un opérateur chargé de les examiner. L'opérateur pouvait accepter l'image telle quelle et l'envoyer directement à la saisie des données par clavier (passant outre la reconnaissance automatisée), ou envoyer le questionnaire entier à l'étape de la vérification à la sortie pour être retiré. Voir ci-dessous pour de plus amples détails sur la saisie des données par clavier et la vérification à la sortie.
- Reconnaissance automatisée : Elle visait à reconnaître automatiquement les réponses écrites à la main et les marques inscrites sur le questionnaire.
- Saisie des données par clavier : Des opérateurs ont entré manuellement les réponses que le système de reconnaissance automatisée ne pouvait pas saisir avec suffisamment de certitude. Environ 13 % de toutes les réponses ont été transmises à la saisie.
- Vérification à la sortie : Après avoir franchi toutes les étapes susmentionnées, les questionnaires papier ont été retirés du système. La vérification à la sortie est un processus d'assurance de la qualité destiné à s'assurer que les images et les données saisies sont d'une qualité telle que les questionnaires papier ne nécessitent pas un traitement ultérieur. Les questionnaires dans lesquels on a décelé des erreurs ont été retirés à la sortie et traités à nouveau.

3.4 Contrôles de couverture, contrôles d'intégralité et suivi des questionnaires rejetés au contrôle

À cette étape, un certain nombre de contrôles automatisés ont été exécutés sur les données des répondants. Ces contrôles ont servi à déceler les cas où le nombre de personnes dénombrées dans le ménage était incorrect en raison d'une erreur de collecte, d'une erreur du répondant ou d'une erreur de saisie des données. La plupart de ces erreurs se sont produites dans les questionnaires papier et comprennent :

- les données entrées par erreur dans la mauvaise colonne relative à la personne
- les données barrées ou rayées saisies par erreur
- les données qui n'ont pas été fournies pour chaque membre du ménage figurant sur la liste au début du questionnaire.

Des erreurs peuvent se produire tant dans les questionnaires en ligne que dans les questionnaires papier; par exemple :

- des données sont fournies pour la même personne dans plus d'un questionnaire (p. ex. une personne remplit son propre questionnaire 3A, mais est également inscrite sur le questionnaire 2A de son ménage)
- des questionnaires sont reçus en double (p. ex. une personne répond au questionnaire sur Internet et son conjoint ou sa conjointe remplit la version papier et l'envoie par la poste).

Environ 58 % des rejets au contrôle ont été résolus automatiquement par le système, lorsque la nature de l'erreur était telle que la solution était évidente. Les solutions consistaient à supprimer les données sur les personnes fictives provenant d'une erreur du répondant ou d'une erreur de saisie, et à supprimer les réponses en double. Les rejets au contrôle n'ayant pas pu être résolus automatiquement ont été transmis aux commis au dépouillement. Un système interactif permettait aux commis de comparer les données saisies d'un questionnaire à l'autre et d'examiner l'image des questionnaires papier pour détecter les erreurs de saisie et les erreurs des répondants. Les rejets au contrôle ont été résolus en supprimant les personnes non valides ou en double ou en ajoutant les personnes manquantes (c.-à-d. en créant des enregistrements de personne en blanc) lorsque nécessaire.

Les contrôles de couverture ont été suivis d'une autre série de contrôles automatisés, afin de déceler les questionnaires où les réponses manquantes étaient trop nombreuses, ou lorsque les données n'avaient pas été fournies pour tous les membres habituels du ménage, y compris les cas où des personnes manquantes avaient été ajoutées par les commis aux contrôles de couverture. Les ménages rejetés au contrôle ont fait l'objet d'un suivi. Un intervieweur téléphonait aux répondants pour régler les problèmes de couverture et obtenir les renseignements manquants, au moyen d'une application d'interview téléphonique assistée par ordinateur. Dans le cas des ménages qui avaient répondu au questionnaire détaillé, seules les données manquantes aux questions qui figuraient également dans le questionnaire abrégé ont fait l'objet d'un suivi. Les données obtenues lors du suivi étaient réintégrées au système en vue d'un traitement subséquent. Dans les cas où le suivi n'a pas permis d'obtenir les données manquantes, les données ont été imputées à l'étape de la vérification et de l'imputation (voir la section 3.9).

3.5 Codage

Les questionnaires du recensement contenaient des questions dont les réponses pouvaient être cochées sur une liste, ainsi que des questions dont les réponses devaient être inscrites par le répondant dans les cases prévues à cette fin. Chaque réponse écrite s'est vu attribuer automatiquement un code numérique au moyen des fichiers de référence, des ensembles de codes et des classifications types de Statistique Canada. Les fichiers de référence utilisés pour le processus d'appariement automatisé ont été créés au moyen des réponses réelles recueillies lors des recensements antérieurs, et à l'aide de fichiers administratifs. Des codeurs et des spécialistes du domaine ayant reçu une formation particulière réglaient les cas où il était impossible d'attribuer automatiquement un code. Les questions suivantes nécessitaient un codage tant dans le questionnaire détaillé que dans le questionnaire abrégé :

- lien avec la Personne 1
- langue parlée à la maison
- langue maternelle.

Les questions suivantes nécessitaient un codage pour l'échantillon du questionnaire détaillé seulement :

- lieu de naissance
- citoyenneté
- langues non officielles
- origine ethnique
- groupe de population
- Première Nation / bande indienne
- lieu de résidence un an auparavant
- lieu de résidence cinq ans auparavant
- lieu de naissance des parents
- principal domaine d'études
- lieu des études
- industrie
- profession
- lieu de travail
- langue de travail.

Un total d'environ 69 millions de réponses écrites ont été codées à partir des questionnaires du Recensement de 2016. Globalement, environ 85 % de ces réponses ont été codées automatiquement, mais le taux de codage automatique variait considérablement d'une variable à l'autre.

3.6 Classification et ajustements pour la non-réponse des logements inoccupés et des logements non répondants

L'Enquête sur la classification des logements (ECL) a servi à estimer le taux d'erreur des agents recenseurs lors de la classification des logements privés comme étant occupés ou inoccupés dans les unités de collecte (UC) où le recensement a été effectué par envoi par la poste et par listage/livraison. L'information ainsi recueillie a permis d'apporter des ajustements à la base de données du recensement. L'ECL a été réalisée auprès d'un échantillon aléatoire de 1 730 UC visées par l'envoi par la poste et le listage/livraison. Les agents recenseurs ont de nouveau visité ces UC en juin et juillet 2016 afin de réévaluer le statut d'occupation le jour du recensement de chaque logement pour lesquels aucune réponse n'avait été reçue. L'ECL a montré que 15,0 % des 1 187 392 logements privés classés comme étant inoccupés étaient en fait occupés, et que 36,9 % des 284 966 logements privés sans réponse qui avaient été classés comme étant occupés ou dont le statut d'occupation était inconnu, étaient en fait inoccupés. Les estimations fondées sur l'échantillon de l'ECL ont servi à ajuster le statut d'occupation des logements individuels. Cela s'est traduit par une augmentation de 2,6 % du nombre de logements privés occupés, et par une diminution de 6,2 % du nombre de logements inoccupés à l'échelle du Canada.

Après l'ajustement du statut d'occupation par l'ECL, les logements privés occupés ayant une non-réponse totale ont fait l'objet d'une imputation du nombre de résidents habituels (s'il n'était pas connu) et de toutes les réponses aux questions du recensement en empruntant les réponses d'un autre ménage répondant dans la même UC. Ce processus, appelé imputation de ménages entiers (IME), a imputé 99,9 % des ménages ayant une non-réponse totale. L'utilisation d'un seul donneur dans le cadre de l'IME était informatiquement plus efficace et moins susceptible de produire des résultats non plausibles que l'utilisation de plusieurs donneurs dans le cadre du processus principal de vérification et d'imputation. Néanmoins, pour les autres ménages ayant une non-réponse totale, c'est-à-dire 0,1 % d'entre eux, pour lesquels aucun ménage donneur n'avait été trouvé dans le cadre de l'IME, l'imputation a été effectuée au moyen du processus principal de vérification et d'imputation.

Le processus de l'IME comporte un autre élément, outre l'utilisation des estimations de l'ECL, pour ajuster la base de données du recensement. Les aires non couvertes par l'ECL — les UC où les questionnaires du recensement sont administrés par des intervieweurs (c.-à-d. les UC dans les réserves indiennes, les UC de recensement par interview et les UC collectives) — nécessitent une autre stratégie d'imputation. Dans ces aires seulement, tous les logements inoccupés sont traités comme étant réellement inoccupés, et tous les logements non répondants sont traités comme étant réellement occupés. Partant du principe que tous les logements inoccupés sont classés correctement, **aucune imputation n'est effectuée**. Les logements privés dont le statut d'occupation est classé comme inconnu sont également traités comme étant inoccupés. En revanche, les logements privés non répondants que les agents recenseurs ont classés comme étant occupés sont tous traités comme s'ils étaient occupés, et le voisin le plus rapproché géographiquement est utilisé comme ménage donneur pour ces logements. Aucune restriction n'a été imposée à la taille du ménage pour ces imputations, contrairement aux aires couvertes par l'ECL. À l'échelle du Canada (tant pour les aires couvertes par l'ECL que celles non couvertes par l'ECL), 2,6 % des logements privés occupés ont été imputés par l'entremise du processus de l'IME.

Des renseignements plus détaillés sur l'ECL et la procédure d'IME seront disponibles dans le *Rapport technique sur la couverture, Recensement de la population, 2016* (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-303/index-fra.cfm), n° 98-303-X au catalogue de Statistique Canada, qui sera publié en 2019.

3.7 Obtention des données sur le revenu

En 2016, pour la première fois, les données administratives ont constitué l'unique source d'information sur le revenu pour le Programme du recensement. Cela a non seulement allégé le fardeau de réponse, mais a aussi accru la qualité et la quantité des données disponibles sur le revenu. Les renseignements sur le revenu des particuliers ont été compilés à partir de données administratives pour l'ensemble de la population âgée de 15 ans et plus plutôt que pour un échantillon, comme c'était le cas en 2011 et 2006. Les revenus imposable et non imposable, de nature régulière et récurrente, reçus pendant l'année civile 2015 ont été inclus. Les revenus ponctuels, comme les retraits forfaitaires de régimes enregistrés d'épargne-retraite et d'autres régimes d'épargne, les règlements forfaitaires d'assurance ou de prestations de retraite, les gains ou pertes en capital, les héritages et les gains de loterie ont été exclus.

L'information sur les répondants au recensement pouvait être couplée à deux types de dossiers de l'Agence du revenu du Canada (ARC) selon qu'ils étaient : 1) des déclarants, pour lesquels toute l'information sur le revenu pouvait être tirée des fichiers d'impôt, soit les déclarations T1, les feuillets fiscaux et les programmes gouvernementaux administrés par l'ARC, ou 2) des non-déclarants, pour lesquels seule l'information provenant des feuillets fiscaux et des programmes gouvernementaux administrés par l'ARC était disponible. En 2016, les renseignements de 94,8 % de la population âgée de 15 ans et plus dans les ménages privés ont été couplés à un dossier administratif de l'ARC. Plus précisément, les renseignements de 85,2 % de la population ont été couplés à un dossier de déclarant, et ceux de 9,6 % de la population ont été couplés à un dossier de non-déclarant.

Pour plus d'information sur l'obtention des données sur le revenu, veuillez consulter le *Guide de référence sur le revenu, Recensement de la population, 2016* (http://www12.statcan.gc.ca/census-recensement/2016/ref/guides/004/98-500-x2016004-fra.cfm), nº 98-500-X2016004 au catalogue.

3.8 Non-réponse

Le statut de non-réponse peut différer à la collecte et au traitement. Les différences principales surviennent parce que le statut d'occupation peut changer entre la collecte et le traitement et parce que le ménage doit répondre à un minimum d'items pour qu'il soit considéré comme répondant à l'étape du traitement. À moins d'avis contraire, lorsqu'il sera mention de « non-réponse », on fera référence à la non-réponse au traitement des données. Il en sera de même si l'on aborde la réponse au lieu de la non-réponse.

Pour le questionnaire détaillé du Recensement de 2016, il y a deux genres de ménages qui sont considérés comme des ménages non répondants :

- les ménages de l'échantillon qui n'ont répondu qu'à des questions communes aux deux types de questionnaires, c'est-à-dire uniquement à des questions du questionnaire abrégé
- les ménages qui n'ont répondu à aucune question.

On parle ici de non-réponse totale. Celle-ci est traitée différemment selon le mode de collecte et le genre de ménage.

Enfin, on parle de non-réponse partielle lorsque le questionnaire détaillé est rempli partiellement. Ce genre de réponse est traité par imputation. Un aperçu de cette méthode est présenté à la section suivante.

3.9 Vérification et imputation

Les données recueillies dans le cadre de tout recensement ou enquête contiennent des omissions ou des incohérences. Par exemple, un répondant peut éprouver de la réticence à répondre à une question, ne pas se souvenir de la bonne réponse ou mal comprendre la question. D'autres erreurs, comme le codage incorrect d'une réponse, peuvent aussi se produire.

L'épuration définitive des données, à l'étape de la vérification et de l'imputation, a été automatisée en utilisant le Système canadien de vérification et d'imputation du recensement (SCANVIR) [Statistique Canada, 2014], pour l'ensemble des sujets du recensement. Deux méthodes d'imputation ont été appliquées. La première méthode, appelée « imputation déterministe », consiste à attribuer des valeurs particulières sous certaines conditions quand la résolution du problème est claire et non ambiguë. Des règles de vérification détaillées sont appliquées pour déterminer ces conditions, puis des valeurs prédéterminées sont attribuées aux variables intervenant dans les règles. La deuxième méthode, appelée « imputation par donneur à changements minimaux par le plus proche voisin », comprend l'application d'une série de règles de vérification détaillées qui permettent de déceler toute réponse manquante ou incohérente. Quand un enregistrement contenant des réponses manquantes ou incohérentes est repéré, un autre enregistrement qui satisfait les règles et dont la plupart des caractéristiques sont les mêmes que celles de l'enregistrement erroné est sélectionné. Les données provenant de cet enregistrement donneur sont empruntées et utilisées pour apporter le nombre minimal de changements aux variables afin de résoudre tous les problèmes de réponses manquantes ou incohérentes.

3.10 Pondération

Le programme du Recensement canadien de 2016 est composé d'un Recensement de la population et d'une enquête-échantillon pour laquelle le quart des ménages privés canadiens sont sélectionnés. Les ménages non échantillonnés pour l'enquête ont reçu un questionnaire abrégé, tandis que les ménages échantillonnés ont reçu un questionnaire détaillé. En plus des questions du questionnaire abrégé, le questionnaire détaillé a permis de recueillir des renseignements socioculturels, des informations sur les activités de la vie quotidienne, la mobilité, le lieu de naissance, la scolarité, les activités sur le marché du travail, etc. On a fait appel à la pondération pour représenter l'ensemble de la population à partir de l'information recueillie auprès de l'échantillon.

La première étape de la pondération consistait à attribuer à chaque ménage un poids de sondage reflétant la probabilité qu'il soit échantillonné. Par la suite, ces poids ont subi un premier ajustement pour tenir compte de la couverture et de la non-réponse totale. Cet ajustement a été appliqué aux poids des ménages répondants. Finalement, un deuxième ajustement, le calage final, a été effectué afin d'établir une meilleure concordance entre les estimations obtenues à partir des ménages répondants de l'échantillon et les chiffres du recensement pour un certain nombre de caractéristiques provenant du questionnaire abrégé ou de sources de données administratives. La méthode de pondération est décrite en détail au chapitre 4. Tous les ménages privés annexés à un logement collectif et tous les ménages privés d'UC de recensement par interview ont été sélectionnés pour l'échantillon du questionnaire détaillé et ont donc reçu un poids de sondage de un. Par la suite, ils ont été exclus des processus d'ajustement pour la couverture et la non-réponse ainsi que du calage final.

Les ménages de l'échantillon du questionnaire détaillé ayant un poids non nul à la fin des étapes de pondération sont les ménages répondants ainsi que les ménages ayant reçu un poids de sondage de un, c.-à-d. les ménages privés annexés à un logement collectif et tous les ménages privés d'UC de recensement par interview. Ces ménages constituent l'ensemble des ménages qui contribuent aux estimations du questionnaire détaillé.

3.11 Taux de réponse finaux

Le tableau 3.11.1 présente les taux de réponse finaux des ménages privés du Recensement de la population de 2016 à l'échelle nationale et pour chaque province et territoire, suivis des taux de réponse non pondérés et pondérés pour l'échantillon du questionnaire détaillé en se basant sur la définition de la non-réponse de la section 3.8.

Le taux de réponse final est calculé par le ratio du numérateur au dénominateur suivant :

- au numérateur : le nombre de logements privés pour lesquels un questionnaire a été rempli[3]
- au dénominateur : le nombre de logements privés classés comme occupés selon la base de données du recensement.

La classification finale du statut d'occupation des logements est fondée sur l'analyse des données recueillies par le personnel sur le terrain, les données fournies par les répondants et les résultats d'une étude de la qualité concernant le statut d'occupation selon l'ECL (voir la section 3.6). Rappelons que les taux de réponse du tableau 3.11.1 se distinguent des taux de réponse à la collecte, auparavant publiés et dont il est fait mention à la section 1.3, en ce sens qu'ils tiennent compte du traitement des données et de la vérification du statut d'occupation des logements pour identifier les ménages non répondants. Ces taux de réponse sont ainsi considérés comme finaux.

3. Les ménages échantillonnés non répondants de Wood Buffalo pour lesquels des données administratives ont été utilisées (voir l'annexe D), de même que tous les logements privés annexés à des logements collectifs qui, par défaut, n'ont reçu qu'un questionnaire abrégé, sont considérés comme ayant rempli leur questionnaire pour les taux de réponse du questionnaire abrégé. Étant donné qu'ils n'ont pas rempli le questionnaire détaillé, ils sont considérés comme non-répondants pour le calcul des taux de réponse du questionnaire détaillé.

Les taux de réponse pondérés sont produits pour l'échantillon du questionnaire détaillé. Ils sont basés sur les poids suivants au numérateur et au dénominateur :

- au numérateur : le poids de sondage
- au dénominateur : les poids calés au total du recensement de la région géographique (ce qui correspond au nombre de ménages privés et occupés selon le recensement).

Tableau 3.11.1
Taux de réponse finaux des ménages privés au Recensement de la population de 2016 et de l'échantillon du questionnaire détaillé

Province/territoire	Taux de réponse du questionnaire abrégé	Taux de réponse non pondéré du questionnaire détaillé seulement	Taux de réponse pondéré du questionnaire détaillé seulement
	pourcentage		
Canada	97,4	96,1	95,9
Terre-Neuve-et-Labrador	97,4	96,0	95,1
Île-du-Prince-Édouard	97,5	96,4	96,3
Nouvelle-Écosse	97,6	96,6	96,1
Nouveau-Brunswick	97,6	96,7	96,2
Québec	97,6	96,7	96,6
Ontario	97,6	96,5	96,3
Manitoba	97,4	95,9	95,8
Saskatchewan	96,7	95,6	95,1
Alberta	97,0	95,3	94,8
Colombie-Britannique	96,5	94,8	94,6
Yukon	95,8	91,9	92,8
Territoires du Nord-Ouest	93,9	92,7	93,0
Nunavut	92,7	92,6	92,6

Note : Tous les ménages privés et occupés sont inclus dans le calcul de ces taux de réponse, sans exception.
Source : Statistique Canada, Recensement de la population, 2016 et échantillon du questionnaire détaillé du Recensement de 2016.

4. Estimation à l'aide de l'échantillon du questionnaire détaillé du recensement

Tout processus d'échantillonnage nécessite une méthode d'estimation connexe pour convertir les données-échantillon à l'échelle de la population et pour s'assurer que les estimations d'après les données d'enquête sont représentatives de la population. Le choix d'une méthode d'estimation est généralement régi par des contraintes de nature opérationnelle ainsi que théorique. Du point de vue opérationnel, la méthode doit être applicable dans le système de traitement des données dont elle fait partie, et du point de vue théorique, la méthode doit réduire au minimum l'erreur statistique dans les estimations produites.

La méthode d'estimation génère un ensemble de poids, et le poids de chaque unité d'échantillonnage correspond au nombre d'unités dans la population que représente l'unité d'échantillonnage. Ces poids sont appliqués aux données-échantillon pour produire des millions d'estimations à l'aide de l'échantillon du questionnaire détaillé du recensement. Les estimations sont des mesures agrégées telles que les totaux, les moyennes, les proportions et les médianes, que l'on calcule à l'aide de l'échantillon pour diverses caractéristiques pertinentes.

4.1 Éléments à considérer pour choisir une méthode d'estimation

4.1.1 Considérations de nature opérationnelle

Mathématiquement, une méthode d'estimation peut être décrite au moyen d'une formule algébrique, ou estimateur, qui montre comment l'estimation pour la population est calculée sous forme d'une fonction des valeurs d'échantillon observées et d'autres données provenant du plan de sondage ou de sources externes. La plupart du temps, cet estimateur est une simple fonction des poids et de la variable d'intérêt pour les unités répondantes. L'utilisation d'un ensemble unique de poids pour produire toutes les estimations garantit un certain niveau de cohérence entre les diverses estimations de l'enquête.

Par conséquent, l'approche adoptée pour l'échantillon du questionnaire détaillé du recensement (et la plupart des enquêtes-échantillon) consistait à subdiviser la méthode d'estimation en deux étapes : a) le calcul des poids (appelé procédure de pondération); b) l'utilisation des poids pour produire les estimations, comme l'estimation d'un chiffre particulier de population par totalisation des poids appliqués aux personnes ou aux ménages possédant la caractéristique pertinente. La complexité mathématique est alors principalement confinée à l'étape a), qui est exécutée une seule fois. L'étape b), quant à elle, est réduite à un simple processus, comme la sommation des poids chaque fois qu'une totalisation est nécessaire. Étant donné que le poids appliqué à chaque unité d'échantillonnage est le même pour toute totalisation faisant intervenir l'unité en question, la cohérence entre les différentes estimations fondées sur les données-échantillon est assurée.

4.1.2 Considérations d'ordre théorique

Pour un plan de sondage donné et une méthode d'estimation donnée, il est possible, en s'appuyant sur la théorie de l'échantillonnage, de formuler un énoncé quant aux chances qu'un certain intervalle contienne la valeur inconnue de la population qu'il faut estimer. Le principal critère relatif au choix d'une méthode d'estimation est la minimisation de l'étendue de ces intervalles pour un niveau donné de confiance, de façon que les énoncés au sujet des valeurs inconnues de la population soient aussi précis que possible. La mesure de précision habituelle pour comparer les méthodes d'estimation est appelée erreur-type. Pourvu que certaines conditions soient respectées, les intervalles de plus ou moins deux erreurs-types par rapport à l'estimation contiendront la vraie valeur de la population pour approximativement 95 % de tous les échantillons possibles.

Outre la minimisation de l'erreur-type, un second objectif relatif au choix d'une méthode d'estimation pour l'échantillon du questionnaire détaillé est de garantir, autant que possible, que les estimations d'après l'échantillon des caractéristiques évaluées au recensement concordent avec les valeurs connues du recensement. Heureusement, ces deux objectifs sont habituellement complémentaires en ce sens que l'erreur d'échantillonnage a tendance à être réduite lorsqu'on veille à ce que les estimations de certaines caractéristiques de base de l'échantillon concordent avec les chiffres de population correspondants. Cependant, bien que cela soit vrai en général, forcer les estimations de l'échantillon du questionnaire détaillé pour des caractéristiques évaluées au

recensement à concorder avec les chiffres du recensement correspondants pour de très petits sous-groupes peut avoir un effet indésirable sur l'erreur-type des estimations pour les caractéristiques de l'échantillon proprement dites. Par exemple, si plusieurs aires de diffusion ne comptent qu'un petit nombre de sujets ayant une certaine caractéristique, par exemple un pays de naissance donné, le fait d'assurer la cohérence entre les estimations de l'échantillon et les chiffres du recensement pour ce pays de naissance ferait augmenter de manière injustifiée l'erreur-type pour le reste des caractéristiques.

Si l'on ne dispose d'aucune information au sujet de la population échantillonnée à part celle recueillie pour les unités d'échantillonnage et en l'absence de non-réponse totale, la méthode d'estimation est limitée à la pondération des unités d'échantillonnage par l'inverse de leurs probabilités de sélection. Par exemple, si une unité avait une chance sur quatre d'être sélectionnée, elle recevrait un poids de 4. Lorsque la non-réponse totale est observée, le poids doit également être ajusté en fonction de l'estimation de la probabilité de réponse de l'unité, par exemple. En pratique, on dispose souvent de certains renseignements supplémentaires au sujet de la population (p. ex. sa taille totale et, parfois, sa répartition en fonction d'une variable donnée — peut-être selon la province ou le territoire). Ce genre d'information peut être utilisé pour améliorer la formule d'estimation afin de produire des estimations ayant plus de chances d'être proches de la valeur inconnue de la population. Dans le cas de l'échantillon du questionnaire détaillé du recensement, on dispose d'un grand nombre de données très détaillées sur la population échantillonnée grâce aux données du questionnaire abrégé du recensement, à tous les niveaux géographiques. Dans le cadre des ajustements pour la couverture, la non-réponse et le calage, on peut profiter de cette quantité considérable de renseignements sur la population pour améliorer les estimations tirées de l'échantillon du questionnaire détaillé.

Néanmoins, on ne peut pas faire concorder les estimations de l'échantillon du questionnaire détaillé pour les caractéristiques du recensement avec tous les chiffres du recensement à chaque niveau géographique. On observe des différences entre les estimations de l'échantillon et les chiffres du recensement lorsqu'on produit un tableau croisé d'une variable-échantillon et de la variable de recensement correspondante. Le tableau des totaux estimés d'après l'échantillon pour des caractéristiques particulières ne concordera pas nécessairement avec le tableau équivalent des totaux des chiffres du recensement pour ces caractéristiques.

L'ajustement, aussi faible que possible, des poids afin d'obtenir une concordance parfaite entre les estimations du questionnaire détaillé et les chiffres du recensement pour certaines caractéristiques et certains sous-groupes porte le nom de « calage ».

4.2 Régions de pondération

Les divers ajustements effectués sur les poids de sondage ont été faits indépendamment selon la région de pondération. Les régions géographiques qui ont été utilisées à cette fin sont les aires de diffusion agrégées (ADA) et les super aires de diffusion agrégées (SADA). L'ADA est une nouvelle région géographique de diffusion créée pour le Recensement de 2016. La SADA est créée spécifiquement pour les procédures de pondération par l'agrégation d'ADA. Il est à noter que les subdivisions géographiques qui ont été utilisées en 2011 et avant étaient construites différemment.

4.2.1 Aire de diffusion agrégée

L'aire de diffusion agrégée (ADA) est une nouvelle région géographique de diffusion créée pour le Recensement de 2016. Son objectif est de permettre la diffusion d'un plus grand nombre de données à un niveau géographique détaillé à l'échelle du pays. Au total, le Canada est divisé en 5 386 ADA, et il y a eu des ménages sélectionnés pour l'échantillon du questionnaire détaillé dans 5 143 ADA. Parmi les 243 ADA où il n'y a pas eu de ménages échantillonnés, 235 étaient formées uniquement de ménages hors du champ de l'enquête. Dans les 8 autres ADA, il n'y avait qu'un petit nombre de ménages dans le champ de l'enquête et aucun d'entre eux n'a été sélectionné.

Les ADA respectent les critères de délimitation suivants :

1. Les ADA couvrent l'ensemble du pays et, dans la mesure du possible, ont une population de 5 000 à 15 000 habitants (selon les chiffres de population du recensement précédent).
2. Les ADA respectent les limites provinciales et territoriales ainsi que les limites des divisions de recensement (DR), des régions métropolitaines de recensement (RMR) et des agglomérations de recensement (AR) comptant des secteurs de recensement (SR) en vigueur au Recensement de 2016.
3. Les ADA sont établies à partir de l'une des trois régions géographiques de diffusion du Recensement de 2016 : les aires de diffusion (AD), les subdivisions de recensement (SDR) ou les secteurs de recensement (SR) :
    - Dans les RMR et les AR comptant des SR, les SR adjacents sont regroupés afin de respecter le critère de population de l'ADA.
    - Dans les régions sans SR (soit les régions à l'extérieur des RMR et des plus grandes AR), où les SDR ont une population inférieure à 15 000 habitants, les SDR adjacentes sont regroupées afin de respecter le critère de population de l'ADA.
    - Dans les régions sans SR, où les SDR ont une population supérieure à 15 000 habitants, les AD adjacentes sont regroupées dans ces SDR afin de respecter le critère de population de l'ADA.
4. Chaque SDR formée d'une réserve indienne et un petit nombre d'autres régions où l'on a eu recours au recensement par interview constituent des ADA distinctes.

« Pour plus de renseignements sur l'Aire de Diffusion Agrégée, consultez le *Dictionnaire, Recensement de la population, 2016*, (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/index-fra.cfm) nº 98-301-X au catalogue. »

Le tableau 4.2.1.1 présente certaines statistiques indiquant dans quelle mesure les ADA comptant des ménages dans l'échantillon du questionnaire détaillé sont bien ajustées aux SDR. Le premier scénario s'est produit dans la grande majorité des cas, car les ADA ont été formées de sorte à respecter en priorité les limites géographiques des SR et des SDR. Le scénario 4 est le seul où les limites géographiques des SDR ne sont pas respectées. Les SR ne figurent pas dans le tableau car tous, sauf un, étaient dans le premier scénario et celui qui n'y était pas se trouvait dans le scénario 3.

Tableau 4.2.1.1
Nombre de subdivisions de recensement qui respectent les limites des ADA comptant des ménages dans l'échantillon du questionnaire détaillé, Recensement de 2016

Scénario	Description	Subdivisions de recensement	
		nombre	pourcentage
1	La SDR était de taille suffisamment petite pour être contenue entièrement dans une ADA, et cette même ADA ne comprenait que des SDR entières. Aucune SDR dans l'ADA ne faisait aussi partie d'une autre ADA.	4 512	92,40
2	La SDR était de taille suffisamment petite pour être contenue entièrement dans une ADA, mais une autre SDR dans la même ADA faisait aussi partie d'une autre ADA.	81	1,66
3	La SDR était de taille suffisamment grande pour contenir des ADA entières. Aucune des ADA ne faisait aussi partie d'une autre SDR.	261	5,35
4	La SDR faisait partie d'au moins deux ADA.	29	0,59
Total		**4 883**	**100,00**

SDR = Subdivision de recensement.
ADA = Aire de diffusion agrégée.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

Le tableau 4.2.1.2 montre la ventilation du nombre d'ADA comptant des ménages dans l'échantillon du questionnaire détaillé selon la province ou le territoire.

Tableau 4.2.1.2
Nombre d'ADA comptant des ménages dans l'échantillon du questionnaire détaillé selon la province ou le territoire

Région	Nombre d'ADA
Terre-Neuve-et-Labrador	81
Île-du-Prince-Édouard	21
Nouvelle-Écosse	146
Nouveau-Brunswick	124
Québec	1 118
Ontario	1 655
Manitoba	216
Saskatchewan	256
Alberta	516
Colombie-Britannique	916
Yukon	28
Territoires du Nord-Ouest	40
Nunavut	26
Canada	**5 143**

ADA = Aire de diffusion agrégée.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

Le tableau 4.2.1.3 montre le nombre d'ADA selon le nombre de ménages du recensement dans le champ de l'enquête. La majorité des ADA comptant des ménages dans l'échantillon du questionnaire détaillé comptent entre 2 000 et 4 999 ménages. Un nombre assez important d'ADA sont peu populeuses.

Tableau 4.2.1.3
Répartition des ADA comptant des ménages dans l'échantillon du questionnaire détaillé selon le nombre de ménages dans le champ de l'enquête

Ménages dans le champ de l'enquête	Nombre d'ADA	Pourcentage
0 à 499	976	18,98
500 à 999	117	2,27
1 000 à 1 999	366	7,12
2 000 à 2 999	1 339	26,04
3 000 à 3 999	1 229	23,90
4 000 à 4 999	664	12,91
5 000 à 5 999	300	5,83
6 000 à 6 999	98	1,91
7 000 à 7 999	32	0,62
8 000 à 8 999	14	0,27
9 000 à 9 999	3	0,06
10 000+	5	0,10
Total	**5 143**	**100,00**

ADA = Aire de diffusion agrégée.
Source : Statistique Canada, Recensement de la population, 2016.

Le tableau 4.2.1.4 présente le nombre d'ADA selon les intervalles du nombre de ménages répondants au questionnaire détaillé du Recensement de 2016. Pour les ADA comptant le moins de répondants, un traitement particulier a été appliqué afin d'avoir suffisamment de ménages pour la pondération (voir la section 4.5). Globalement, les ADA comptaient davantage de répondants que les régions de pondération utilisées en 2011.

Tableau 4.2.1.4
Répartition des ADA comptant des ménages dans l'échantillon du questionnaire détaillé selon le nombre de ménages répondants au questionnaire détaillé

Ménages répondants	Nombre d'ADA	Pourcentage
0 à 99	605	11,76
100 à 199	265	5,15
200 à 299	158	3,07
300 à 399	142	2,76
400 à 499	311	6,05
500 à 599	537	10,44
600 à 699	628	12,21
700 à 799	604	11,74
800 à 899	518	10,07
900 à 999	400	7,78
1 000 à 1 099	304	5,91
1 100 à 1 199	229	4,45
1 200 à 1 299	162	3,15
1 300 à 1 399	101	1,96
1 400 à 1 499	73	1,42
1 500 et plus	106	2,06
Total	**5 143**	**100,00**

ADA = Aire de diffusion agrégée.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

4.2.2 Super aire de diffusion agrégée

Les super aires de diffusion agrégées ont été créées spécifiquement pour la pondération du Recensement de 2016 afin de mener certaines procédures de pondération pour lesquelles il est souhaitable d'avoir un grand nombre d'observations.

Les SADA ont été établies de sorte à respecter les règles suivantes (en ordre de priorité) :

1. Les SADA sont créées en regroupant des ADA (obligatoire).
2. Les SADA respectent les limites provinciales et territoriales (obligatoire).
3. Les SADA ont une population de 50 000 à 150 000 personnes (sauf pour les divisions de recensement (DR) qui ont une population de 40 000 à 50 000 personnes qui elles forment leur propre SADA.
4. La population des SADA exclue les personnes vivant dans des UC de recensement par interview.
5. Les SADA respectent les limites des divisions de recensement (DR).
6. Les SADA respectent les limites des régions métropolitaines de recensement (RMR) et des agglomérations de recensement (AR).
7. Les SADA respectent les limites des subdivisions de recensement.
8. Les SADA sont en un seul morceau, donc contigües.
9. Les SADA sont aussi compactes que possible.

Les deux premières règles étaient obligatoires et les règles 3 à 9 ont été respectées autant que possible. Au total, 409 SADA ont été formées.

Le tableau 4.2.2.1 montre la ventilation du nombre de SADA selon la province ou le territoire.

Tableau 4.2.2.1
Nombre de SADA selon la province ou le territoire

Région	Nombre de SADA
Terre-Neuve-et-Labrador	8
Île-du-Prince-Édouard	2
Nouvelle-Écosse	13
Nouveau-Brunswick	8
Québec	97
Ontario	150
Manitoba	15
Saskatchewan	14
Alberta	44
Colombie-Britannique	55
Yukon	1
Territoires du Nord-Ouest	1
Nunavut	1
Canada	**409**

SADA = Super aire de diffusion agrégée.
Note : Pour les trois territoires, la SADA correspond au territoire.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

Le tableau 4.2.2.2 présente certaines statistiques indiquant dans quelle mesure les SADA sont bien ajustées aux DR et aux RMR. Les SADA respectent les limites géographiques de la grande majorité des DR (scénarios 1 et 3) et respectent celles des trois quarts des RMR. Les autres RMR font partie d'au moins deux SADA (scénario 4).

Tableau 4.2.2.2
Nombre de divisions de recensement et de régions métropolitaines de recensement qui respectent les limites des SADA, Recensement de 2016

		Divisions de recensement		Régions métropolitaines de recensement	
Scénario	**Description**	nombre	pourcentage	nombre	pourcentage
1	La DR ou la RMR était de taille suffisamment petite pour être contenue entièrement dans une SADA, et la même SADA ne comprenait que des DR ou des RMR entières. Aucune DR ni aucune RMR dans la SADA ne faisait aussi partie d'une autre SADA.	249	84,98	2	5,71
2	La DR ou la RMR était de taille suffisamment petite pour être contenue entièrement dans une SADA, mais une autre DR ou RMR dans la même SADA faisait aussi partie d'une autre SADA.	2	0,68	0	0,00
3	La DR ou la RMR était de taille suffisamment grande pour contenir des SADA entières. Aucune des SADA ne faisait aussi partie d'une autre DR ou RMR.	40	13,65	25	71,43
4	La DR ou la RMR faisait partie d'au moins deux SADA.	2	0,68	8	22,86
Total		**293**	**100,00**	**35**	**100,00**

DR = Division de recensement.
RMR = Région métropolitaine de recensement.
SADA = Super aire de diffusion agrégée.
Source : Statistique Canada, Recensement de la population, 2016.

Le tableau 4.2.2.3 montre le nombre de SADA selon le nombre de personnes dans le champ de l'enquête.

Tableau 4.2.2.3
Répartition des SADA comptant des ménages dans l'échantillon du questionnaire détaillé selon le nombre de personnes dans le champ de l'enquête

Personnes dans le champ de l'enquête	Nombre de SADA	Pourcentage
30 000 à 39 999	2	0,49
40 000 à 49 999	26	6,36
50 000 à 59 999	23	5,62
60 000 à 69 999	45	11,00
70 000 à 79 999	106	25,92
80 000 à 89 999	67	16,38
90 000 à 99 999	46	11,25
100 000 à 149 999	94	22,98
Total	**409**	**100,00**

SADA = Super aire de diffusion agrégée.
Source : Statistique Canada, Recensement de la population, 2016.

4.3 Poids de sondage

Le poids de sondage de chaque ménage de l'échantillon du questionnaire détaillé a été calculé différemment selon le mode de collecte de l'UC dans lequel se trouve le logement correspondant.

Si le mode de collecte est :

- l'envoi par la poste, le poids de sondage correspond à l'inverse de la fraction de sondage, ce qui correspond à un poids de quatre
- le listage/livraison, le poids de sondage correspond au ratio du nombre de logement privés dénombrés au nombre de logement privés échantillonnés dans l'UC et ces poids sont d'environ quatre pour 98 % des ménages sélectionnés, et ils varient de un à sept
- le recensement par interview, le poids de sondage est de un.

Les ménages vivant dans des logements privés annexés à un logement collectif font exception. Comme mentionné à la section 2.2, tous ces ménages ont été ajoutés à l'échantillon. Ils sont considérés à tirage complet et ont par le fait même reçu un poids de sondage de un.

4.3.1 Poids des ménages dénombrés dans l'échantillon

Les ménages échantillonnés qui ont un poids de sondage de un ne reçoivent aucun ajustement de poids. Ces ménages conservent leur poids de un une fois les procédures de pondération terminées (couverture et non-réponse, ainsi que calage aux totaux du recensement). Ils proviennent d'UC de recensement par interview ou bien sont des ménages privés annexés à des logements collectifs.

Pour ces ménages, la non-réponse totale et la non-réponse partielle sont traitées par imputation. Une fois les données manquantes imputées, ces ménages sont considérés comme répondants aux fins d'estimation (bien qu'ils soient non-répondants pour le calcul des taux de réponse présentés à la section 3.11).

4.4 Ajustement pour la couverture et pour la non-réponse totale

Les diverses manières de traiter la non-réponse aux enquêtes peuvent être classées en deux catégories principales : l'imputation et la repondération. La première est généralement employée pour traiter les valeurs manquantes isolées. La deuxième sert surtout pour les cas de non-réponse totale. Un ménage était traité comme un répondant au questionnaire détaillé lorsqu'il répondait à au moins une question du questionnaire détaillé. Étant donné le taux de réponse élevé au questionnaire détaillé, toute méthode d'ajustement aux cas de non-réponse aurait eu un effet généralement moindre sur les estimations et les poids définitifs de l'enquête. La couverture et la non-réponse totale des ménages dans les UC situées dans les réserves indiennes et dans les UC de recensement par interview ont été compensées par des procédures d'imputation, principalement par l'imputation de ménages entiers (IME) telle que décrite dans la section 3.6. Dans le reste du pays, des procédures de repondération ont été employées. Le reste du présent chapitre décrit ces procédures de pondération.

Les ajustements pour la couverture et pour la non-réponse visent principalement à réduire l'incidence de tout biais causé par une couverture incomplète (ou des questionnaires reçus en double) ou par la non-réponse totale. Pour que l'ajustement réduise réellement le biais potentiel, il est très utile d'avoir un vaste ensemble de renseignements sur les non-répondants. En l'absence de tels renseignements, l'ajustement pour la non-réponse que l'on peut appliquer sera limité et ne pourra pas réduire de beaucoup le biais potentiel. On connaissait les renseignements géographiques de tous les ménages non répondants et de tous les non-répondants de l'échantillon du questionnaire détaillé (c.-à-d. les répondants de l'échantillon du questionnaire détaillé qui ont répondu aux questions du questionnaire abrégé, mais pas à celles du questionnaire détaillé). Les renseignements sur les non-répondants étaient donc plutôt limités. Heureusement, avant que ne soient appliqués les ajustements pour la couverture et pour la non-réponse, le processus d'IME a été effectué. Une part importante de l'IME consistait à imputer les caractéristiques du questionnaire abrégé pour tous les non-répondants à ce questionnaire. Cela

comprenait les non-répondants de l'échantillon du questionnaire détaillé. Ces renseignements supplémentaires ont servi de base à l'ajustement pour la non-réponse pour l'échantillon du questionnaire détaillé.

Aux fins d'ajustement pour la couverture et pour la non-réponse totale dans l'échantillon du questionnaire détaillé, on a appliqué une procédure de repondération fondée sur le calage aux poids de sondage. Cette procédure peut être divisée en quatre étapes principales :

1. sélection des contraintes de calage pour les étapes 2 et 3
2. ajustement pour la couverture selon un calage non linéaire
3. estimation de la propension à la non-réponse en fonction du calage non linéaire de la non-réponse
4. application d'une méthode de pointage fondée sur la propension à l'étape 3.

Les étapes 1 à 4 ont été appliquées de manière indépendante dans chaque SADA. En d'autres termes, l'ajustement pour la non-réponse a été appliqué par SADA. La section 4.2 présente des définitions et des renseignements ayant trait aux ADA et aux SADA.

La première étape consistait en une sélection ascendante des contraintes de calage dans la SADA, qui a été effectuée comme suit :

- L'ensemble de contraintes potentielles a été tiré des variables communes au questionnaire abrégé et au questionnaire détaillé, ainsi que de données administratives obtenues au moyen de stratégies de couplage des enregistrements (au cours desquelles toutes les unités de la population du questionnaire détaillé sont soumises aux procédures de couplage). En raison des exigences de la méthode de calage non linéaire utilisée à la deuxième et à la troisième étape, seules les contraintes au niveau de la SADA ainsi que le nombre de ménages et de personnes dans chaque ADA de la SADA ont été pris en considération.
- Dans chaque SADA, deux contraintes obligatoires ont été sélectionnées en premier : le nombre de ménages dans la SADA (TOTHHLD) et le nombre de personnes dans la SADA (TOTPERS).
- Les contraintes propres aux ADA — le nombre de ménages (HHADA) et le nombre de personnes (PPADA) — ont été évaluées pour la sélection.
- Toutes les autres contraintes potentielles de la SADA ont été évaluées, en donnant priorité à celles qui divisaient le plus possible la population de la SADA en deux parts égales.

Le processus de sélection excluait les contraintes qui visaient moins de 250 ménages dans la SADA et les contraintes jugées redondantes ou presque redondantes à ces contraintes ou à des contraintes déjà sélectionnées. Les contraintes redondantes à des contraintes déjà sélectionnées ont été exclues parce qu'elles n'apportaient aucun nouveau renseignement. Compte tenu de ces filtres, l'ordre de priorité employé dans l'évaluation des contraintes a fait en sorte que les contraintes sélectionnées se complétaient mutuellement et compensaient tout écart potentiel de couverture entre le questionnaire détaillé et le questionnaire abrégé, ainsi que pour la non-réponse totale au recensement.

La deuxième étape consistait à appliquer un ajustement pour la couverture selon un calage non linéaire à l'échantillon entier dans la SADA (c.-à-d. les répondants comme les non-répondants). Pour les contraintes sélectionnées à la première étape, les comptes pondérés de l'échantillon du questionnaire détaillé ont été harmonisés avec les chiffres de population correspondants. Cette étape avait pour but de corriger tout écart potentiel de la couverture entre l'échantillon du questionnaire détaillé et son complément (c.-à-d. l'ensemble des ménages recevant uniquement le questionnaire abrégé). Le surdénombrement peut se produire notamment lorsque des personnes sont comptées dans deux ménages différents. La couverture pour les deux populations pourrait également être différente si, par exemple, les logements occupés étaient plus susceptibles d'être incorrectement répertoriés comme étant inoccupés selon le questionnaire détaillé que selon le questionnaire abrégé. Un autre objectif de cette étape consistait à isoler autant que possible l'erreur d'échantillonnage. Sans cette étape, le calage de la non-réponse à l'étape suivante confondrait l'erreur due à la non-réponse avec l'erreur d'échantillonnage. Cette étape fait concorder les estimations de l'échantillon avec les estimations de population.

En outre, les mêmes totaux de contrôle sont utilisés dans les deux procédures de calage. Par conséquent, l'estimation de la propension à la non-réponse effectuée à l'étape suivante n'a pas besoin de corriger (directement ou indirectement) l'erreur d'échantillonnage. Combiner une correction pour l'erreur d'échantillonnage et pour l'erreur due à la non-réponse à l'étape suivante aurait été inapproprié. La procédure de calage aurait échoué si le poids de tout répondant avait dû être réduit en vue de concorder avec les chiffres du recensement, car la propension estimée aurait alors dépassé 1. De plus, la méthode de pointage appliquée à la dernière étape nécessitait une estimation de la propension de réponse toute seule. Dans la mesure où la variable d'intérêt était liée aux contraintes sélectionnées, la variance d'échantillonnage a également été réduite lors de cette étape.

Après ces deux étapes, le principal ajustement pour la non-réponse a été effectué. Le poids des non-répondants, ajusté à l'étape précédente, a été mis à 0 tandis que le poids des répondants a été augmenté pour faire en sorte que les sommes pondérées de la SADA concordent avec les chiffres de population correspondants pour les contraintes sélectionnées. Une fonction de lien logistique entre la propension de réponse et les caractéristiques utilisées lors du calage a permis une estimation implicite de la propension de réponse. Folsom et Singh (2000) ont proposé cette méthode de calage non linéaire comme méthode pour ajuster la non-réponse tout en s'assurant que les deux estimations correspondaient aux chiffres de population sélectionnés et que les probabilités de réponse estimées se trouvaient entre 0 et 1. Cette dernière condition ne reste pas nécessairement valable lorsqu'un calage linéaire est utilisé pour l'ajustement pour la non-réponse. Dans la mesure où la propension de réponse était liée aux contraintes sélectionnées, le biais de non-réponse potentiel a été réduit sans augmenter la variance lors de cette étape.

L'inverse des probabilités de réponse estimées obtenues lors de l'étape précédente pourrait être utilisé directement pour ajuster le poids de la non-réponse. Cependant, la méthode de pointage a été utilisée pour la dernière étape de l'ajustement pour la non-réponse afin d'« aplanir » les probabilités estimées de l'étape précédente. Cela a permis d'assurer la qualité de l'ajustement pour la non-réponse et d'éviter les ajustements trop extrêmes. Pour chaque ADA, des classes de pondération homogènes ont été constituées en fonction des probabilités de réponse estimées. Dans chaque classe, la moyenne harmonique pondérée des probabilités de réponse a été calculée. La moyenne harmonique a été utilisée parce qu'elle est moins affectée par les valeurs aberrantes des probabilités de réponse estimées. L'inverse de cette moyenne a été appliqué au poids des répondants de la classe à titre d'ajustement pour la non-réponse.

En bref, l'ajustement pour la couverture et pour la non-réponse totale était un produit de deux quantités : l'ajustement pour la couverture et l'inverse de la moyenne harmonique obtenue selon la méthode de pointage.

4.5 Calage final

Le calage final est un calage linéaire et il a été effectué dans le but de minimiser la variabilité d'échantillonnage des estimations dérivées des réponses au questionnaire détaillé tout en assurant la cohérence entre les totaux estimés et les totaux du Recensement de la population. Cette étape de pondération était nécessaire, car il était important d'assurer la cohérence entre les totaux estimés et les totaux du Recensement de la population pour un grand nombre de variables et de régions géographiques, c.-à-d. de respecter des contraintes de calage.

Seuls les poids des ménages dont le mode de collecte de l'UC était l'envoi par la poste ou le listage/livraison ont été calés, car ces ménages ont été échantillonnés. Les ménages faisant exception à cette règle sont ceux de ces UC, mais provenant de logements privés annexés à un logement collectif. Parce que tous ces ménages font partie de l'échantillon du questionnaire détaillé et que toutes les réponses au questionnaire détaillé de ces ménages ont été imputées, aucun calage n'a été effectué. Les poids finaux de ces ménages sont donc égaux à un. Les poids produits par le processus de calage sont les poids finaux servant à calculer les estimations provenant du questionnaire détaillé et ces poids s'appliquent autant aux ménages qu'aux familles et aux personnes, c.-à-d. que toutes les familles et personnes d'un même ménage reçoivent le poids du ménage. Pour ce dernier ajustement, il était important de limiter la variabilité des poids calés afin d'éviter qu'une part excessive de poids ne soit appliquée à un ménage ou à une personne. Par conséquent, les poids ont été contraints de prendre une valeur comprise entre 1 et 20.

Les contraintes de calage ont été définies aux niveaux de la personne, du ménage et de la famille de recensement. De plus, en 2016, on a ajouté deux niveaux géographiques hiérarchiques dans la notion de contraintes, soit les niveaux d'ADA et de SADA. L'ajout de ces deux niveaux a permis de maximiser de façon globale la cohérence entre les totaux estimés et les totaux du Recensement de la population tout en minimisant le nombre de contraintes de calage, ce qui devrait permettre de réduire la variabilité des estimations. L'annexe C énumère toutes les contraintes qui ont été prises en considération durant le processus de calage au niveau des ADA et au niveau des SADA. Les caractéristiques pour lesquelles des données provenant du recensement, de sources administratives et du questionnaire détaillé étaient disponibles et pour lesquelles on s'est efforcé d'établir une concordance comprennent, entre autres, l'âge, le sexe, l'état matrimonial, l'union libre, la taille du ménage, le type de logement et la langue officielle parlée.

Le choix des contraintes s'est fait dans les deux types de régions géographiques simultanément et de façon indépendante. Par la suite, le calage a été exécuté en utilisant toutes les contraintes sélectionnées. En 2016, l'ajout de contraintes de calage à deux niveaux géographiques (ADA et SADA), le retrait des contraintes de langue maternelle et l'ajout de contraintes basées sur des données administratives appariées signifient que le nombre de contraintes potentielles était différent du Recensement de 2011. Au niveau des ADA, 270 contraintes ont été définies, tandis que 200 contraintes ont été définies au niveau des SADA. Différentes raisons ont justifié le choix du niveau géographique des contraintes de calage. Ce choix s'est fait en collaboration avec les spécialistes des domaines spécialisés. Par exemple, certaines contraintes n'ont été définies qu'au niveau des SADA, car elles n'auraient pas été assez populeuses au niveau des ADA. Pour d'autres contraintes telles que les groupes d'âge, en plus de vouloir des contraintes assez populeuses, le choix s'est fait de façon à ce que les contraintes ne soient pas trop semblables l'une de l'autre lorsqu'elles sont évaluées par la procédure de sélection.

Pour faciliter le calage des petites ADA, celles-ci ont été regroupées avant d'effectuer la sélection des contraintes de calage afin d'avoir un minimum de 60 ménages répondants au questionnaire détaillé par ADA. Les petites ADA appartenant entièrement à une SDR ont premièrement été regroupées avec d'autres ADA de la même SADA. Par la suite, les petites ADA des DR ont été regroupées avec d'autres ADA de la même SADA. Enfin, les petites ADA restantes ont été regroupées à une ADA appartenant à une SADA voisine. La procédure de groupement des ADA a produit 4 180 groupements d'ADA de 60 ménages répondants ou plus.

La première étape du processus de sélection des contraintes de calage consistait à classer les contraintes en trois groupes :

Contraintes obligatoires : Ces contraintes doivent être utilisées dans le calage parce qu'il doit y avoir concordance entre les chiffres du recensement et les estimations provenant du questionnaire détaillé aux niveaux géographiques qui sont des agrégats habituels des ADA et des SADA (p. ex. Canada, provinces et territoires). Le nombre de personnes et le nombre de ménages dans les ADA et les SADA étaient les deux contraintes obligatoires.

Contraintes à faible réponse : Les contraintes évaluées pour une population de 200 ménages ou moins ne sont pas utilisées dans le calage parce qu'elles peuvent rendre les estimations de l'enquête instables.

Toutes les autres contraintes : Ces contraintes ont été examinées plus en détail afin de déterminer s'il fallait les utiliser dans le calage.

La deuxième étape consistait à déterminer quelles contraintes du troisième groupe devaient être utilisées dans le processus de calage en plus des contraintes obligatoires. Les contraintes de ce troisième groupe ont été ajoutées une par une en choisissant itérativement la contrainte divisant le plus également possible la population de la SADA ou de l'ADA en deux. Les contraintes qui étaient trop linéairement dépendantes ont été exclues. De plus, afin d'éviter d'introduire un biais dans les estimations ponctuelles et d'éviter d'augmenter leur variance, le nombre de contraintes sélectionnées a été limité. À la suite d'évaluations, il a donc été déterminé que ce nombre devait être plus petit que la racine carrée du nombre de ménages répondants participant à la contrainte.

Après avoir sélectionné les contraintes de calage à utiliser, une dernière vérification a été faite afin de vérifier que l'ensemble des contraintes choisi au niveau des ADA et au niveau de la SADA soit exempt de colinéarité.

Le calage proprement dit a ensuite été exécuté pour l'ensemble final de contraintes de la deuxième étape en modifiant les poids ajustés pour la couverture et la non-réponse aussi peu que possible de manière à ce que les estimations pondérées soient égales aux totaux du recensement pour ces contraintes. Le Système généralisé d'estimation (SGE) de Statistique Canada a été utilisé pour effectuer le calage.

Il existe quelques raisons pour lesquelles les estimations de l'échantillon peuvent différer des chiffres du recensement, particulièrement pour les petites régions, même après l'étape du calage. Quelques-unes de ces raisons sont énumérées ci-dessous.

- Contraintes exclues durant la sélection des contraintes : Comme il est décrit plus haut, des contraintes potentielles pouvaient être exclues parce que le nombre de réponses était faible, parce qu'elles étaient linéairement dépendantes à d'autres contraintes choisies (ou trop dépendantes) ou parce qu'elles étaient linéairement dépendantes à des contraintes pour lesquelles le nombre de réponses était faible (ou trop dépendantes). Cela a donné lieu à des différences entre les chiffres du recensement et les estimations provenant du questionnaire détaillé pour ces variables lorsqu'il ne s'agissait pas de dépendance linéaire parfaite avec des contraintes choisies.
- Sous-région de pondération : En 2016, l'ADA était la plus petite région de pondération pour laquelle on a essayé d'obtenir une concordance entre les chiffres du recensement et les estimations provenant du questionnaire détaillé. Toute entité plus petite qu'une ADA, comme la plupart des AD, est appelée sous-région de pondération. Des écarts pouvaient exister entre les chiffres du recensement et les estimations provenant du questionnaire détaillé pour ces sous-régions de pondération.

4.6 Détails sur la sélection des contraintes[4]

Au cours du processus de pondération, la sélection de contraintes a été effectuée à deux reprises : lors de l'ajustement pour la couverture et la non-réponse, qui fait appel à des techniques de calage non linéaire, et lors du calage final. Les variables constituant les contraintes étaient essentiellement les mêmes, mais l'inclusion ou l'exclusion de contraintes variaient entre les deux étapes de pondération, car leurs objectifs respectifs étaient différents. En résumé, la sélection des contraintes ne fait pas exactement appel aux mêmes critères, et les régions de pondération varient selon l'étape de pondération.

La présente section explique comment la sélection des contraintes est effectuée de part et d'autre de ces étapes de pondération et donne la fréquence à laquelle certaines contraintes ont été exclues durant l'ajustement pour la couverture et la non-réponse de même qu'au calage final. La sélection des contraintes a été effectuée indépendamment pour les 408 SADA qui comportent des ménages échantillonnés dont le poids est ajusté, ainsi que pour les 4 238 ADA qui les composent. Une seule SADA n'a pas subi d'ajustement de poids car elle ne contenait que des ménages dénombrés (la SADA correspondant au Nunavut). Cela explique pourquoi les tableaux qui suivent présentent les résultats de 408 SADA alors que le nombre total de SADA figurant à la section 4.2 est de 409. Durant l'ajustement pour la couverture et la non-réponse, 4 646 régions de pondération ont été définies, soit 408 SADA et 4 238 ADA, tandis que 4 588 régions de pondération ont été définies à l'étape du calage final. Le nombre de régions de pondération diffère selon l'étape de pondération car à l'étape du calage final, les petites ADA ont été combinées en « ADA de calage » de façon à avoir un minimum de 60 ménages par ADA, ce qui a donné un total de 4 180 ADA à l'étape du calage.

4. La sélection des contraintes spécifiques à la région de Wood Buffalo fait exception, puisque cette région a fait l'objet d'un processus particulier. Pour cette raison, la sélection des contraintes se rapportant à cette subdivision de recensement est exclue de ce chapitre pour être présentée à l'annexe D.

4.6.1 Processus d'ajustement pour la couverture et la non-réponse

La procédure d'ajustement pour la couverture et la non-réponse utilise le calage pour ajuster les poids de sondage. La justification est que si les estimations fondées sur les répondants concordent autant que possible avec les chiffres du recensement pour les variables auxiliaires, le biais de non-réponse des estimations associées à ces variables sera réduit. La section 4.4 décrit plus en détail l'ajustement pour la couverture et la non-réponse.

Comme le mentionne la section 4.4, l'ajustement pour la couverture et la non-réponse a été apporté de manière indépendante dans la SADA, où plusieurs variables ont été utilisées pour définir les contraintes, et dans l'ADA, pour deux variables. Une procédure séquentielle a été utilisée pour sélectionner les contraintes. Dans chaque SADA, deux contraintes obligatoires ont été sélectionnées en premier : le nombre de ménages dans la SADA (TOTHHLD) et le nombre de personnes dans la SADA (TOTPERS). Par la suite, des contraintes propres aux ADA — le nombre de ménages (HHADA) et le nombre de personnes (PPADA) — dans chacune des ADA ont été prises en considération. Ensuite, toutes les autres contraintes potentielles de la SADA ont été prises en considération, en privilégiant celles qui divisaient (approximativement) la SADA en deux parts égales. Les contraintes non obligatoires ont été ajoutées de manière séquentielle à l'ajustement, à condition de ne pas correspondre aux critères cités dans le tableau 4.6.3.1. La liste définitive des contraintes de calage qui ont été prises en considération pour l'ajustement pour la couverture et la non-réponse se trouve à l'annexe C.

Une contrainte qui est fréquemment exclue présente généralement un plus grand écart entre le chiffre du recensement et l'estimation de l'échantillon ajusté pour la non-réponse qu'une contrainte qui est utilisée plus souvent. Cela est observable lorsque l'on compare l'annexe C avec le tableau 5.2.1. L'annexe C présente toutes les variables ou les contraintes potentielles, le nombre de fois qu'une contrainte a été utilisée pour le calage, et le nombre de fois que les contraintes ont été exclues pour l'une des raisons indiquées au tableau 4.6.3.1. Cinq critères ont pu mener à l'exclusion d'une contrainte de l'ajustement pour la couverture et la non-réponse : « Population nulle », « Faible population », « Dépendance linéaire », « Grande colinéarité » et « Redondance explicative ». Il est possible qu'une contrainte soit exclue du calage pour l'une de ces raisons, mais qu'elle soit calée à la fin du processus. Cela se produit par exemple lorsque la contrainte est colinéaire avec les contraintes sélectionnées. Dans un tel cas, l'annexe C montre cette contrainte comme étant « calée ».

4.6.2 Procédures pour l'ajustement du calage final

L'objectif du calage final est d'ajuster les poids des ménages afin que les estimations produites à partir du questionnaire détaillé soient aussi proches que possible des chiffres du recensement pour de nombreuses caractéristiques communes. En plus de produire une concordance entre les estimations, un choix judicieux de contraintes aura pour effet de réduire la variance. L'annexe C fournit la liste complète des contraintes potentielles et la section 4.5 donne une description du calage effectué sur les estimations provenant du questionnaire détaillé. Les critères appliqués pour la sélection des contraintes sont similaires à ceux appliqués à la sélection des contraintes pour l'ajustement de la couverture et de la non-réponse, avec quelques variantes telles que présentées au tableau 4.6.3.1.

Le calage a été effectué simultanément pour les SADA et les ADA de calage. Dans chacune des SADA, des contraintes de calage ont été définies au niveau de la SADA et au niveau des ADA. Dans chaque SADA, toutes les contraintes ont été évaluées et n'ont été exclues qu'au besoin. Le nombre total de personnes (TOTPERS) et le nombre total de ménages (TOTHHLD) étaient les deux seules contraintes obligatoires, ce qui signifie qu'elles ne pouvaient être exclues pour aucune des ADA.

À l'échelle canadienne, un total de 132 777 contraintes présélectionnées a été envoyé au SGE développé par Statistique Canada afin que celui-ci effectue le calage linéaire final, ce qui représente une moyenne de 27 contraintes par ADA et une moyenne de 31 contraintes par SADA. Les contraintes obligatoires ont été sélectionnées dans toutes les régions de pondération. Le processus mis en place a fait en sorte que les contraintes de revenu, en particulier celle du revenu du ménage ainsi que celle indiquant un ménage à faible revenu, ont été sélectionnées le plus souvent. Les contraintes sélectionnées le moins souvent correspondent principalement à l'année d'immigration ainsi qu'au pays d'origine.

4.6.3 Comparaison des procédures entre les deux étapes d'ajustement et sélection

À chacune des étapes de pondération, des critères ont été appliqués lors de la sélection des contraintes. Ces critères sont présentés au tableau 4.6.3.1 selon l'étape de pondération.

Tableau 4.6.3.1
Critères appliqués lors de la sélection des contraintes pour l'ajustement de la couverture, de la non-réponse et du calage final

Critère	Ajustement de la couverture et de la non-réponse	Calage final
Population nulle selon les chiffres du recensement : Si la contrainte ne comportait aucune population dans l'aire de pondération, l'estimation après ajustement doit aussi être égale à 0 pour cette contrainte. Ces contraintes ne sont pas classées comme étant exclues, mais sont plutôt inadmissibles au processus d'ajustement.	Appliqué au niveau de la SADA/ADA.	Appliqué au niveau de la SADA/ADA.
Petite population selon les chiffres du recensement : Si une contrainte fait intervenir moins qu'un certain nombre de ménages dans la population de l'aire de pondération, on considère que cette contrainte est négligeable et on l'exclut. Inclure ce type de contrainte augmenterait exagérément la variance. Toutefois, les contraintes assorties d'une petite population peuvent être implicitement calées et, dans ce cas, on les inclut dans le nombre total des contraintes calées.	Appliqué au niveau de la SADA/ADA. Le nombre de ménages dans la population de la région de pondération est supérieur à 0 mais inférieur à 250.	Appliqué au niveau de la SADA/ADA. Le nombre de ménages dans la population de la région de pondération est supérieur à 0 mais inférieur à 200.
Linéairement dépendant : S'il est possible de calculer la valeur d'une contrainte en combinant les valeurs d'autres contraintes, l'une de ces contraintes n'est pas nécessaire et elle doit être supprimée lors du processus d'ajustement en raison de sa dépendance linéaire. Toutefois, les contraintes exclues en raison de leur dépendance linéaire sont implicitement calées. Celles-ci sont donc incluses dans le total du nombre de contraintes calées.	Appliqué au niveau de la SADA. La sélection des contraintes peut être comparée avec la sélection des variables explicatives dans une régression linéaire. Par conséquent, on utilise le facteur d'inflation de la variance[1] (FIV) et l'indice de conditionnement[2] pour détecter une colinéarité élevée.	Appliqué au niveau de la SADA/ADA. Deux vérifications de dépendance sont effectuées pour identifier les contraintes linéairement dépendantes : une première lors de la sélection des contraintes au niveau de la SADA/ADA) et une deuxième sur l'ensemble complet des contraintes choisies aux deux niveaux géographiques hiérarchiques (SADA et ADA qui la composent).

Tableau 4.6.3.1
Critères appliqués lors de la sélection des contraintes pour l'ajustement de la couverture, de la non-réponse et du calage final

Critère	Ajustement de la couverture et de la non-réponse	Calage final
Colinéarité élevée : Si la valeur d'une contrainte peut quasiment être calculée en la combinant avec d'autres valeurs de contrainte, il conviendra de supprimer l'une de ces contraintes lors du processus d'ajustement. Une contrainte de ce type n'est pas parfaitement calée.	Appliqué au niveau de la SADA. La sélection des contraintes peut être comparée avec la sélection des variables explicatives dans une régression linéaire. Par conséquent, on utilise le facteur d'inflation de la variance[1] (FIV) et l'indice de conditionnement[2] pour détecter une colinéarité élevée.	Appliqué au niveau de la SADA/ADA. Deux vérifications de dépendance linéaire sont effectuées pour identifier les contraintes presque linéairement dépendantes : une première lors de la sélection des contraintes au niveau de la SADA et au niveau de la ADA et une deuxième sur l'ensemble complet des contraintes choisies aux deux niveaux hiérarchiques (SADA et ADA qui la composent).
Redondance explicative : Si une contrainte explique la non-réponse (presque) au même titre que d'autres contraintes déjà sélectionnées, la procédure de calage des non-réponses échouerait. Autrement dit, si une contrainte n'ajoute rien de plus aux explications fournies par les contraintes déjà sélectionnées au sujet du mécanisme de non-réponse, il ne faudrait pas l'inclure.	Appliqué au niveau de la SADA. On applique une procédure séquentielle (un genre de régression logistique) pour tester la convergence de la régression logistique.	S/O

1. Le FIV quantifie l'accroissement de la variance des coefficients de régression en raison de la colinéarité.
2. L'indice de conditionnement quantifie le degré selon lequel une matrice est proche de la singularité.
SADA = Super aire de diffusion agrégée.
ADA = Aire de diffusion agrégée.

L'annexe C indique le statut de chacune des contraintes ayant été sélectionnées dans au moins une des régions géographiques une fois la sélection des contraintes effectuée selon l'étape de pondération. La colonne ayant trait à la géographie indique le niveau géographique auquel la contrainte s'appliquait. Dans le cas d'une contrainte s'appliquant aux deux niveaux géographiques, les totaux englobent ces deux niveaux sans différenciation. Il est à noter qu'une contrainte peut avoir été exclue du processus de calage, mais être tout de même calée. Dans ce cas, cette contrainte est considérée comme calée.

4.6.4 Analyse du calage lors de l'ajustement pour la couverture et la non-réponse

La présente section résume le nombre de contraintes qui ont été sélectionnées ou exclues, ainsi que les raisons pour lesquelles certaines contraintes n'ont pas été sélectionnées. En outre, la section explique pourquoi certaines contraintes sont fréquemment exclues.

Les personnes nées dans certains lieux ont tendance à se concentrer davantage dans certaines régions du pays, au point où de nombreuses SADA comptent peu de personnes ayant un lieu de naissance donné ou n'en comptent pas. Par conséquent, les contraintes relatives au lieu de naissance n'ont souvent pas été sélectionnées, en raison des faibles chiffres du recensement. De même, la contrainte relative au français, langue officielle (OLN_FR) compte peu de personnes dans certaines régions du pays ou n'en compte pas, ce qui a souvent mené à son exclusion suppression.

Les contraintes les plus souvent exclues pour cause de grande colinéarité ont été les suivantes :

- « Personnes dans l'ADA » (PPADA)
- « Enfants faisant partie d'une famille de recensement » (CHILD)
- « Familles de recensement avec enfants » (CHILDFAM)
- « Ménages de 1 personne » (HHSIZE1)
- « Femmes » (FEMALE)
- « Hommes » (MALE)
- « Familles de recensement sans enfant » (NOCLDFAM)
- « Femmes âgées de moins de 15 ans » (FEMALELT15)
- « Personnes dans une famille économique » (INEFAM)
- « Personnes dans un ménage ne faisant pas partie d'une famille économique » (NOINEFAM)
- « Hommes âgés de 14 ans et moins » (MALELT15)
- « Personnes âgées de 10 à 14 ans » (AGE14)
- « Ménages de 5 personnes et plus » (HHSIZEGE5)
- « Personnes ne faisant pas partie d'une famille de recensement » (NOTINFAM)
- « Personnes âgées de 5 à 9 ans » (AGE9)
- « Personnes âgées de 0 à 4 ans » (AGE4)
- « Ménages de 6 personnes et plus » (HHSIZEGE6)
- « Adultes faisant partie d'une famille de recensement » (ADULTCF)
- « Personnes faisant partie d'un couple (mariage, partenaire) » (COUPLE)
- « Ménages de 2 personnes » (HHSIZE2).

La procédure a exclue ces contraintes automatiquement, car elles pouvaient être déterminées avec une grande précision à l'aide d'une combinaison des contraintes obligatoires « Ménages » (TOTHHLD) et « Personnes » (TOTPERS) avec d'autres contraintes qui ont souvent été sélectionnées, telles que celles relatives à l'âge, à l'état matrimonial, à la taille du ménage, au sexe selon l'âge et aux personnes dans la famille de recensement. Ces contraintes auraient également pu être expliquées trop facilement par une combinaison des variables sélectionnées et des petites contraintes.

Les différences réelles entre les chiffres du recensement et les estimations ajustées pour la non-réponse sont examinées dans la section 5.2.

Le tableau 4.6.4.1 montre le nombre de fois où chaque raison de supprimer ou d'éliminer une contrainte au niveau de la SADA est survenue. Le nombre total de contraintes exclues est la somme des catégories « Faible population », « Grande colinéarité » et « Redondance explicative ». La catégorie « Population nulle » n'est pas incluse dans le total parce qu'elle ne représente pas réellement des contraintes exclues. Le nombre moyen de contraintes exclues par SADA est le total de chaque catégorie divisé par 408, soit le nombre de SADA où ont été effectués des ajustements pour la couverture et la non-réponse.

Tableau 4.6.4.1
Statistiques sommaires sur les contraintes relatives aux SADA dans l'ajustement pour la couverture et la non-réponse en 2016

Contrainte	Calée	Population nulle	Exclues			Total exclues
			Petite population	Colinéarité élevée	Redondance explicative	
Nombre total de contraintes	30 328	2 865	29 742	11 550	49	41 341
Nombre moyen de contraintes par SADA	74,3	7,0	72,9	28,3	0,1	101,3

SADA = Super aire de diffusion agrégée.
Note : Deux SADA sont exclues, la SADA de Wood Buffalo et une deuxième parce qu'elle est formée de ménages provenant de réserves seulement.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

En moyenne, 74,3 contraintes au niveau de la SADA ont été calées par SADA. Une moyenne de 72,9 contraintes ont été exclues par SADA en raison d'une faible population et 28,3, en raison d'une grande colinéarité.

4.6.5 Analyse du calage final

Sauf dans le cas où la population est nulle, chaque fois qu'une contrainte est exclue le processus de calage n'essaiera pas de faire concorder les estimations du questionnaire détaillé aux chiffres du recensement pour cette contrainte dans la région de pondération en question. L'écart entre le chiffre du recensement et l'estimation provenant du questionnaire détaillé sera habituellement plus grand pour une contrainte exclue fréquemment que pour une contrainte exclue moins souvent.

Le tableau 4.6.5.1 montre à quelle fréquence une contrainte a été exclue ou éliminée au niveau de la région de pondération selon chaque critère. Le nombre total de contraintes exclues est égal à la somme des valeurs pour les différents critères de suppression. Le nombre moyen de contraintes exclues par région de pondération est simplement égal au total pour la catégorie divisé par le nombre de régions de pondération.

Tableau 4.6.5.1
Statistiques sommaires relatives au statut de sélection des contraintes au niveau de la région de pondération dans l'ajustement final des poids en 2016

Contraintes	Région de pondération	Calées	Population nulle	Exclues		Total exclues	Nombre de régions de pondération
				Petite population	Colinéarité élevée		
	ADA	210 307	112 442	283 947	99 319	383 266	4 180
Nombre de contraintes	SADA	39 262	2 698	27 292	10 588	37 880	408
Nombre moyen de contraintes par SADA/ADA	ADA	50,3	27,1	67,9	23,8	91,7	4 180
	SADA	96,2	6,6	66,9	26,0	92,0	408

SADA = Super aire de diffusion agrégée.
ADA = Aire de diffusion agrégée.
Note : Deux SADA sont exclues, la SADA de Wood Buffalo et une deuxième parce qu'elle est formée de ménages provenant de réserves seulement.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

5. Évaluation des procédures de pondération

Comme il est décrit au chapitre 4, la première étape de la pondération de l'échantillon du questionnaire détaillé était l'attribution de poids de sondage aux ménages. L'attribution des poids s'est faite différemment selon le mode de collecte du secteur dans lequel se trouvait le ménage. En effet, seuls les ménages privés ayant un poids de sondage supérieur à un subissent des ajustements. Ceux-ci correspondent aux ménages privés non annexés à un logement collectif provenant d'UC dont le mode de collecte est l'envoi par la poste ou le listage/livraison. Tous les résultats présentés dans le reste du présent chapitre sont calculés pour ce sous-ensemble de ménages. Les poids finaux des autres ménages privés correspondent donc aux poids de sondage et conservent leur valeur initiale de un.

En résumé, rappelons qu'on a attribué à chaque ménage un poids de sondage déterminé par le plan de sondage du questionnaire détaillé. Certains ajustements ont ensuite été nécessaires pour tenir compte de la couverture et de la non-réponse totale. Un calage non linéaire a été effectué lors de cet ajustement pour estimer les paramètres des modèles de non-réponse. Les poids, ajustés pour la couverture et la non-réponse totale, ont par la suite été ajustés durant le processus de calage final afin de produire les poids finaux. Les poids finaux ont permis d'obtenir une concordance généralement meilleure entre les chiffres du recensement et les estimations correspondantes provenant du questionnaire détaillé.

Dans les prochaines sections, on examine la répartition des poids et, pour diverses caractéristiques, les écarts entre les chiffres du recensement et les estimations d'échantillon selon divers niveaux d'agrégation géographique.

5.1 Répartition des poids

Le graphique 5.1.1 et les tableaux 5.1.1 et 5.1.2 montrent la répartition des poids de sondage, des poids ajustés pour la couverture et la non-réponse et des poids ajustés lors du calage final. Les poids sont groupés par intervalles de longueur de 0,5, sauf pour les premier et dernier intervalles. Le graphique montre le pourcentage de fois que les poids se trouvent dans chaque intervalle. On constate que les poids de sondage étaient presque tous compris entre 3,75 et 4,25. Cela est dû au plan de sondage utilisé pour tirer l'échantillon du questionnaire détaillé, selon lequel environ un ménage sur quatre a reçu un questionnaire détaillé dans la plupart des régions. Il est également possible de constater l'effet qu'ont eu les ajustements pour la couverture et la non-réponse ainsi que pour le calage final sur les poids. Une différence très évidente est que le pourcentage de ménages dont le poids de sondage était compris entre 3,75 et 4,25 a été réduit considérablement pour passer d'environ 100 % à 67 %, et terminer à près de 30 %. Les poids finaux étaient répartis plus uniformément entre les catégories comprises entre 1 et 20.

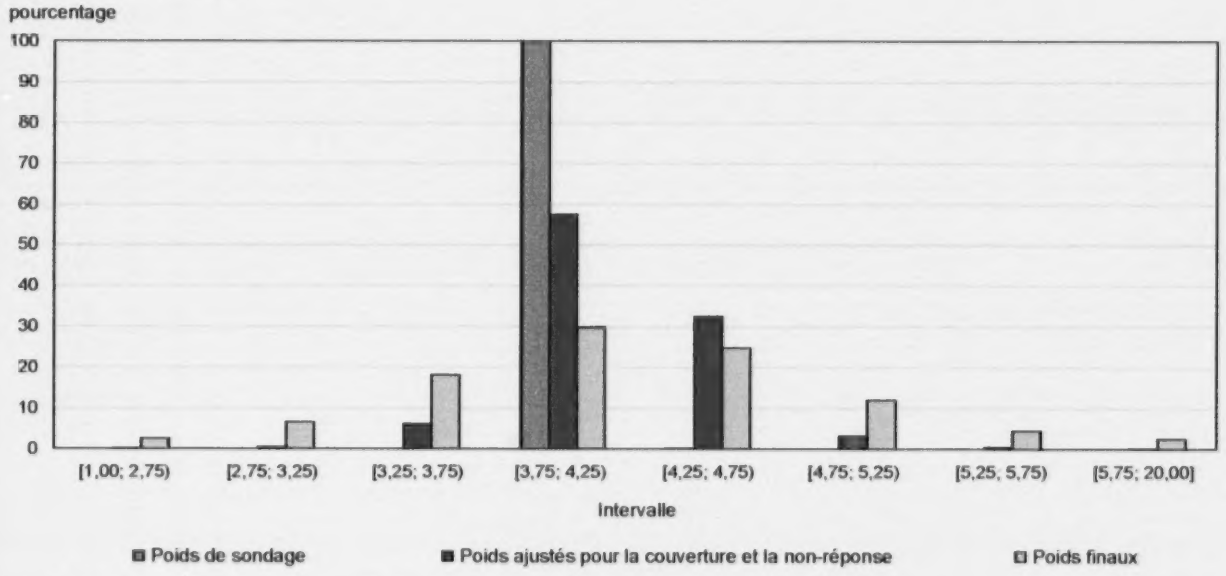
Logiquement, le processus d'ajustement pour la non-réponse aurait tendance à accroître les poids pour compenser pour les unités non répondantes. Cet accroissement s'est produit dans la plupart des cas. Les changements entre les poids de sondage et les poids ajustés pour la couverture et la non-réponse sont décrits dans le tableau 5.1.1. Ce tableau montre que la plupart des unités ayant quitté l'intervalle [3,75 à 4,25) se sont déplacées vers les intervalles [4,25 à 4,75) ou [4,75 à 5,25). Cependant, le processus d'ajustement pour la couverture et la non-réponse a déplacé les poids de l'intervalle 3,75+ aux intervalles [1 à 2,75), [2,75 à 3,25) ou [3,25 à 3,75) pour quelques unités. La raison principale est que la procédure comprenait un ajustement pour le surdénombrement et le sous-dénombrement. Lorsque certains groupes de population ont été surdénombrés, les poids ont été réduits dans ces aires.

Un élément important à noter dans le tableau 5.1.1 est le fait que les unités non répondantes avaient à l'origine un poids positif, puisqu'elles avaient été sélectionnées pour l'échantillon. Le processus d'ajustement pour la non-réponse leur a attribué un poids de 0 et a redistribué leur poids original parmi les unités répondantes. Les unités non répondantes correspondent à la ligne « 0 (non-répondants) » dans le tableau 5.1.1 et ont été exclues du tableau 5.1.2, puisqu'elles n'ont pas été utilisées dans le processus de calage. Le tableau 5.1.2 présente les changements entre les poids ajustés pour la couverture et la non-réponse et les poids calés.

Selon le tableau 5.1.2, la plupart des poids n'ont connu qu'une modification mineure lors du processus de calage. En fait, 76,6 % des cas sont demeurés dans le même intervalle ou n'ont monté ou baissé que d'un seul intervalle. L'intervalle le plus stable était celui situé entre 1 et 2,75 : en effet, 64 % des ménages dont le poids ajusté pour la non-réponse se trouvait dans cet intervalle y sont demeurés après le calage. La deuxième catégorie la plus stable était 5,75+ : 61,8 % des ménages dont le poids ajusté pour la non-réponse était situé entre 5,75 et 9,5 sont demeurés dans cette catégorie (même si l'intervalle de calage va jusqu'à 20, et non jusqu'à 9,5).

Enfin, tandis que les poids ajustés pour la non-réponse ont varié de 1 à 9,5, les poids finaux sont allés de 1 à 20. Cette variabilité des poids est nettement inférieure à la variabilité des poids finaux dans l'Enquête nationale auprès des ménages de 2011 (voir Statistique Canada, 2015), dont le poids maximal avait atteint la limite de 100.

Graphique 5.1.1
Répartition des poids de sondage, des poids ajustés pour la couverture et la non-réponse, et des poids finaux



Notes : Tous les ménages ayant un poids de sondage de 1 ont été exclus du processus de pondération. Ces ménages se trouvaient dans une unité de collecte située dans une réserve indienne ou une unité de collecte de recensement par interview, ou étaient des ménages privés annexés à un logement collectif.
Le symbole « [» signifie que le chiffre est inclus dans l'intervalle, et le symbole «) » signifie que le chiffre n'est pas inclus dans l'intervalle.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

Tableau 5.1.1
Répartition des poids de sondage et des poids ajustés pour la couverture et la non-réponse

Poids ajustés pour la couverture et la non-réponse	Poids de sondage		
	[3,75; 4,25)	[4,25; 7,00]	Total
0 (poids des non-répondants)	122 807	19	**122 826**
[1,00; 2,75)	1 548	0	**1 548**
[2,75; 3,25)	12 363	0	**12 363**
[3,25; 3,75)	201 534	2	**201 536**
[3,75; 4,25)	1 921 456	60	**1 921 516**
[4,25; 4,75)	1 082 895	239	**1 083 134**
[4,75; 5,25)	103 892	85	**103 977**
[5,25; 5,75)	12 783	14	**12 797**
[5,75; 20.00]	3 730	5	**3 735**
Total	**3 463 008**	**424**	**3 463 432**

Notes : Tous les ménages dont le poids de sondage est égal à 1 sont exclus du processus de pondération. Ces ménages se trouvent soit dans une réserve indienne et dans les UC de recensement par interview, ou correspondent à des ménages privés occupant un logement attenant à la structure d'un logement collectif.
Le symbole « [» signifie que le chiffre est inclus dans l'intervalle, et le symbole «) » signifie que le chiffre n'est pas inclus dans l'intervalle.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

Tableau 5.1.2
Répartition des poids finaux et des poids ajustés pour la couverture et la non-réponse

Poids finaux	Poids ajustés pour la couverture et la non-réponse								
	[1,00; 2,75)	[2,75; 3,25)	[3,25; 3,75)	[3,75; 4,25)	[4,25; 4,75)	[4,75; 5,25)	[5,25; 5,75)	[5,75; 9,50)	Total
[1,00; 2,75)	990	3 750	18 082	42 843	12 597	1 402	196	71	79 931
[2,75; 3,25)	340	3 805	38 693	138 180	33 513	2 196	205	37	216 969
[3,25; 3,75)	142	2 910	63 351	414 026	115 038	5 456	454	86	601 463
[3,75; 4,25)	42	1 266	50 385	649 869	275 758	12 551	878	135	990 884
[4,25; 4,75)	27	413	21 470	444 950	334 973	22 686	1 522	219	826 260
[4,75; 5,25)	7	151	6 648	165 518	196 585	25 532	2 354	364	397 159
[5,25; 5,75)	0	48	1 986	47 211	75 313	17 880	2 601	514	145 553
[5,75; 20,00]	0	20	921	18 919	39 357	16 274	4 587	2 309	82 387
Total	**1 548**	**12 363**	**201 536**	**1 921 516**	**1 083 134**	**103 977**	**12 797**	**3 735**	**3 340 606**

Notes : Tous les ménages dont le poids de sondage est égal à 1 sont exclus du processus de pondération. Ces ménages se trouvent soit sur une réserve indienne et dans les UC de recensement par interview, ou correspondent à des ménages privés occupant un logement attenant à la structure d'un logement collectif.
Le symbole « [» signifie que le chiffre est inclus dans l'intervalle, et le symbole «) » signifie que le chiffre n'est pas inclus dans l'intervalle.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

5.2 Écarts entre les chiffres du recensement et les estimations pour le questionnaire détaillé, Canada

Le chapitre 4 décrit les méthodes utilisées pour calculer les poids finaux des ménages et la section 5.1 montre certaines relations entre les poids de sondage, les poids ajustés pour la non-réponse et les poids finaux. L'ajustement pour la couverture et la non-réponse a réduit les écarts entre les chiffres du recensement et les estimations correspondantes pour le questionnaire détaillé pour les contraintes prises en considération (voir l'annexe C). À la suite de ces ajustements, le calage a davantage réduit, voire éliminé ces écarts pour certaines variables (contraintes). Cependant, certains écarts persistent, puisque des contraintes sont parfois exclues (voir

la section 4.6). La différence relative entre les chiffres du recensement et les estimations pour le questionnaire détaillé aussi appelé écart est défini comme suit :

$$\text{Écart} = \frac{(\text{estimation pour le questionnaire détaillé} - \text{chiffre du recensement})}{\text{chiffre du recensement}} \times 100\ \%$$

Le numérateur de l'expression ci-dessus (estimation pour le questionnaire détaillé – chiffre du recensement) désigne la différence entre les chiffres du recensement et les estimations pour le questionnaire détaillé. En divisant cette valeur par le chiffre du recensement, on obtient la différence entre les chiffres du recensement et les estimations pour le questionnaire détaillé par rapport à la grandeur du chiffre du recensement. Autrement dit, le ratio indique le pourcentage associé à la caractéristique qui a été surestimée (valeur positive) ou sous-estimée (valeur négative).

Le tableau 5.2.1 montre les différences entre les chiffres du recensement et les estimations pour le questionnaire détaillé à l'échelle du Canada pour les contraintes prises en considération pour les poids de sondage, les poids ajustés pour la couverture et la non-réponse, et les poids finaux. Le chiffre du recensement pour la caractéristique du nombre total de personnes (TOTPERS) est inférieur au chiffre publié du Recensement de 2016 (35 151 728). Cette différence est attribuable au processus de pondération, qui a uniquement inclus les ménages privés (et non ceux vivant dans des logements annexés à des logements collectifs) et qui a exclu les UC dans des réserves indiennes et les UC de recensement par interview afin d'utiliser une population commune aux fins de la comparaison.

Le tableau 5.2.1 montre également les différences relatives en pourcentage (appelées écarts) pour les estimations en fonction des poids finaux. L'analyse de ces écarts permet de mieux comprendre les différences. Plus de 88 % des cas dans le tableau 5.2.1 avaient un écart entre -1 % et 1 %, et plus de 97 % avaient un écart entre -5 % et 5 %.

Le graphique 5.2.2 montre, pour toutes les contraintes, la différence entre les chiffres du recensement et chacune des trois estimations : selon le poids de sondage (en rouge), ajustée pour la non-réponse (en violet) et selon le poids final (en bleu). L'axe des X représente la taille de la population de la contrainte. Les trois lignes montrent, pour chaque contrainte :

- la différence entre la somme des poids de sondage et le chiffre du recensement
- la somme des poids ajustés pour la non-réponse et le chiffre du recensement
- la somme du poids final et le chiffre du recensement.

Les contraintes sont classées de gauche à droite par ordre croissant de la taille de la population.

Le graphique 5.2.3 montre les écarts en pourcentage entre les chiffres du recensement et les estimations finales de toutes les contraintes selon la taille de la population. Pour les contraintes moyennes ou les grandes contraintes, les écarts sont tous faibles. Seules certaines petites contraintes ont des écarts relativement importants. Par exemple, la différence entre le chiffre du recensement et l'estimation du poids de sondage pour la contrainte « Personne célibataire, jamais mariée » (COMMONLAWNO_SINGLE), selon une taille de la population de 13 624 243, est d'environ -171 000 (rouge). Par contre, l'ajustement pour la non-réponse réduit cette différence à environ -6 000 (violet) seulement, et le calage final réduit encore davantage la différence à environ -2 000 (bleu).

L'observation la plus importante à tirer du graphique 5.2.2 est le fait que l'ajustement pour la couverture et la non-réponse améliore grandement les estimations selon le poids de sondage. La graphique ne l'illustre pas de manière évidente, mais c'est l'ajustement pour la couverture qui joue le plus grand rôle. La différence entre les chiffres du recensement et les estimations pour le questionnaire détaillé pour les poids de sondage avait tendance à être (beaucoup) plus grande que la différence entre les chiffres du recensement et les estimations du questionnaire détaillé pour les poids ajustés pour la non-réponse. Cette dernière différence, pour sa part, avait tendance à être plus grande que la différence correspondante utilisant les poids finaux. Cela illustre l'importance des processus

d'ajustement pour la non-réponse et de calage. Comme on l'a indiqué dans la section 4.6, une différence entre le chiffre du recensement et l'estimation pour le questionnaire détaillé pourrait avoir lieu dans une SADA ou une ADA pour une caractéristique si la contrainte est exclue durant le calage. Cela indique que le processus n'a pas exercé un contrôle sur la contrainte exclue pour une aire donnée. Si la contrainte est exclue dans de nombreuses aires, les différences pourraient s'annuler partiellement, ou elles pourraient s'additionner pour créer une différence plus importante à l'échelle du Canada. Le nombre total de personnes (TOTPERS) et le nombre total de ménages (TOTHHLD) étaient les deux seules contraintes obligatoires pour lesquelles la concordance entre les chiffres du recensement et les estimations pour le questionnaire détaillé devait être garantie pour toutes les ADA. Par conséquent, la différence ainsi que l'écart pour les poids finaux pour ces caractéristiques étaient nuls. Néanmoins, toutes les autres contraintes ont dû être exclues pour certaines aires.

La section 4.6 a indiqué quelles contraintes avaient fréquemment été exclues et où les différences ou les écarts élevés pourraient se trouver. Les effets de la suppression de ces contraintes sont présentés dans le tableau 5.2.1. Voici les contraintes visant uniquement les SADA ayant été exclues le plus fréquemment[5] :

- « Personnes dans une famille économique » (INEFAM)
- « Personnes dans un ménage ne faisant pas partie d'une famille économique » (NOINEFAM)
- « Personnes faisant partie d'un couple (mariage, partenaire) » (COUPLE)
- « Personnes provenant d'un ménage de plus qu'une personne ne faisant pas partie d'une famille économique » (NOINEFAMHHSIZEGT1)
- « Adultes faisant partie d'une famille de recensement » (ADULTCF)
- « Personnes ne faisant pas partie d'une famille de recensement » (NOTINFAM)
- « Enfants faisant partie d'une famille de recensement » (CHILD)
- « Personnes dans un ménage de plus d'une personne ne faisant pas partie d'une famille de recensement » (NOTINFAMHHSIZEGT1)
- « Parents seuls » (LONEPAR).

Il n'est guère surprenant que les différences les plus grandes (celles de plus de 2 000 en valeur absolue) aient été obtenues précisément pour ces contraintes. La plus grande différence a été obtenue pour la contrainte « Personnes dans une famille économique » (INEFAM) (14 493) et son complément, « Personnes dans un ménage ne faisant pas partie d'une famille économique » (NOINEFAM). Cependant, étant donné les chiffres élevés du recensement, les écarts en pourcentage pour toutes ces contraintes n'étaient pas si élevés, allant de -0,93 % à 0,06 %. Dans l'Enquête auprès des ménages de 2011, la différence maximale était de 166 801 (voir Statistique Canada, 2015).

Les écarts les plus grands en pourcentage ont été observés pour certains des lieux de naissance. Comme de nombreux lieux de naissance ne représentent pas un nombre élevé dans bien des parties du pays, ces variables ont souvent été exclues durant le calage. Cela a donné lieu à certaines différences entre les chiffres du recensement et les estimations pour le questionnaire détaillé. En outre, comme le chiffre du recensement est relativement faible pour bon nombre de ces lieux de naissance, les différences relatives sont amplifiées. Par conséquent, les écarts en pourcentage étaient plus importants. L'écart en pourcentage le plus important, soit 12,87 %, a été observé pour la contrainte POBG2_49, qui correspond à « Mongolie ». Cette contrainte a rarement été sélectionnée, car elle comptait une population faible ou nulle dans l'aire. Les autres lieux de naissance pour lesquels les écarts étaient importants — et qui étaient rarement sélectionnés — étaient le « Botswana, Lesotho et Swaziland » (POBG2_9), ayant enregistré un écart de 10,53 %; le « Canada » (POBG2_13), un écart de 8,56 %; « Malte » (POBG2_44), un écart de 6,22 %; les régions « Guadeloupe, Martinique et Saint-Martin (partie française) » (POBG2_15), un écart de 6,1 %; « Andorre, Espagne et Gibraltar » (POBG2_2), un écart de 4,75 %; la « Birmanie (Myanmar) » (POBG2_12), un écart de 4,53 %.

5. Exclues dans plus de 60 % des SADA.

Toutes les caractéristiques des ménages et des familles avaient de faibles écarts en pourcentage; l'écart le plus élevé était une sous-estimation de 0,34 % pour les ménages de 6 personnes et plus (HHSIZEGE6). La plus grande différence pour ce type de caractéristique a été observée du côté des ménages de taille 1, pour lesquels on a enregistré une sous-estimation de 2 025 ménages. D'autres caractéristiques des ménages ayant des différences relativement grandes[6] (mais des écarts néanmoins faibles) étaient « ménages de 2 personnes » (HHSIZE2), « familles de recensement sans enfants » (NOCLDFAM), « familles de recensement » (TOTCFAM), « ménages de 6 personnes et plus » (HHSIZEGE6), et « ménages de 5 personnes et plus » (HHSIZEGE5). L'estimation la plus précise pour une contrainte non obligatoire de ce type appartenait aux « Ménages vivant dans un logement individuel non attenant » (SNGLDET), une différence de 56 et un écart de 0,001 % ayant été observés.

Outre les lieux de naissance, les caractéristiques liées à la personne qui présentaient les plus grands écarts étaient les contraintes « Personnes provenant d'un ménage de plus qu'une personne ne faisant pas partie d'une famille économique » (NOINEFAMHHSIZEGT1), pour laquelle on a enregistré un écart de -0,93 %, et « Personnes dans un ménage de plus d'une personne ne faisant pas partie d'une famille de recensement » (NOTINFAMHHSIZEGT1), pour laquelle on a observé un écart de -0,32 %. Outre celles mentionnées plus haut, voici les caractéristiques relatives à la personne ayant été surestimées ou sous-estimées par une marge d'au moins 1 000 personnes :

- « Personnes célibataires (jamais mariées) et ne vivant pas en union libre » (COMLAWNO_SINGLE)
- « Personnes célibataires âgées de moins de 15 ans (jamais mariées) et ne vivant pas en union libre » (COMLAWNO_SINGLE_LT15)
- « Personnes vivant en union libre » (COMLAW_YE)
- « Personnes mariées » (MARRIED)
- « Pays de naissance - Bélarus, Moldova, Fédération de Russie et Ukraine » (POBG2_28).

Cependant, les écarts pour toutes ces caractéristiques étaient faibles, le plus important n'étant que de 0,74 % pour la dernière caractéristique.

Les estimations les plus précises pour les contraintes non obligatoires liées à la personne étaient pour les personnes des catégories d'âge 35 à 44 ans (AGE35_44) et 45 à 54 ans (AGE45_54) ainsi que pour les « Personnes dont le revenu d'emploi annuel est au dessus du 75e percentile de la SADA » (EMPIN_SADA_P100). Ces caractéristiques ont été estimées sans erreur d'échantillonnage à l'échelle du Canada. Voici d'autres contraintes relatives à la personne ayant présenté de très faibles différences et écarts[7] :

- « Personnes âgées de 55 à 59 ans » (AGE59)
- « Personnes dont le revenu d'emploi annuel est au-dessus du 50e percentile de la SADA », (EMPIN_SADA_GT50)
- « Personnes dont le revenu d'emploi annuel est au-dessus du 50e percentile et inférieur ou égal au 75e percentile de la SADA » (EMPIN_SADA_P75)
- « Personnes âgées de 20 à 24 ans » (AGE24)
- « Personnes dont le revenu d'emploi annuel est égal ou inférieur au 50e percentile de la SADA » (EMPIN_SADA_LE50)
- « Personnes âgées de 15 à 29 ans » (AGE15_29)
- « Personnes âgées de 50 à 64 ans » (AGE50_64).

6. Plus de 1 200 en valeur absolue.
7. Différence inférieure à 10 en valeur absolue et écart inférieur à 0,005 %.

Enfin, parmi les contraintes appliquées uniquement aux SADA, la différence la plus grande (-640) et l'écart en pourcentage le plus grand (-0,02 %) ont été obtenus pour la contrainte « Ménages dont le revenu annuel est inférieur ou égal au 25e percentile de la SADA » (HHINC_SADA_P25). Toutes ces contraintes ont été sélectionnées relativement souvent. La contrainte sélectionnée le moins souvent, « Personnes dont le revenu d'emploi annuel est au-dessus du 25e percentile et inférieur ou égal au 50e percentile de la SADA » (EMPIN_SADA_P50), a été sélectionnée dans 62,7 % des ADA (sous le code EMPIN_P50).

Tableau 5.2.1
Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

		Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		
Caractéristique	Recensement chiffre	estimation	différence	estimation	différence	estimation	différence	écart (%)
ADULTCF	17 873 064	17 796 322	-76 742	17 880 659	7 595	17 879 104	6 040	0,03
AGE00_14	5 679 559	5 629 183	-50 376	5 684 239	4 680	5 679 522	-37	0,00
AGE4	1 845 178	1 828 386	-16 792	1 848 647	3 469	1 845 517	339	0,02
AGE9	1 962 324	1 945 126	-17 198	1 962 857	533	1 962 244	-80	0,00
AGE14	1 872 057	1 855 671	-16 386	1 872 734	677	1 871 761	-296	-0,02
AGE15_24	4 155 227	4 098 212	-57 015	4 152 363	-2 864	4 155 209	-18	0,00
AGE15_29	6 387 808	6 297 713	-90 095	6 383 708	-4 100	6 387 805	-3	0,00
AGE19	1 970 121	1 949 762	-20 359	1 970 676	555	1 970 110	-11	0,00
AGE24	2 185 106	2 148 450	-36 656	2 181 687	-3 419	2 185 099	-7	0,00
AGE25_34	4 513 475	4 453 702	-59 773	4 512 230	-1 245	4 513 547	72	0,00
AGE29	2 232 581	2 199 501	-33 080	2 231 345	-1 236	2 232 597	16	0,00
AGE30_49	9 047 387	8 960 912	-86 475	9 045 676	-1 711	9 047 405	18	0,00
AGE34	2 280 894	2 254 201	-26 693	2 280 885	-9	2 280 950	56	0,00
AGE35_44	4 454 400	4 412 465	-41 935	4 452 867	-1 533	4 454 400	0	0,00
AGE39	2 243 631	2 223 020	-20 611	2 242 432	-1 199	2 243 735	104	0,00
AGE44	2 210 769	2 189 445	-21 324	2 210 436	-333	2 210 665	-104	0,00
AGE45_54	4 936 819	4 905 005	-31 814	4 936 661	-158	4 936 819	0	0,00
AGE49	2 312 093	2 294 247	-17 846	2 311 923	-170	2 312 055	-38	0,00
AGE50_64	7 436 076	7 406 957	-29 120	7 435 705	-371	7 436 084	8	0,00
AGE54	2 624 726	2 610 758	-13 968	2 624 738	12	2 624 764	38	0,00
AGE55_64	4 811 350	4 796 199	-15 151	4 810 967	-383	4 811 321	-29	0,00
AGE59	2 568 136	2 559 441	-8 695	2 569 510	1 374	2 568 127	-9	0,00
AGE64	2 243 214	2 236 757	-6 457	2 241 457	-1 757	2 243 194	-20	0,00
AGE65PL	5 443 624	5 445 343	1 719	5 444 038	414	5 443 637	13	0,00
AGE74	3 302 938	3 302 147	-791	3 303 259	321	3 302 967	29	0,00
AGE75PL	2 140 686	2 143 195	2 509	2 140 779	93	2 140 670	-16	0,00
APTLT5	2 533 749	2 515 260	-18 489	2 532 535	-1 214	2 533 675	-74	0,00
CHILD	10 082 408	9 998 775	-83 633	10 102 364	19 956	10 085 159	2 751	0,03
CHILDFAM	5 722 912	5 681 084	-41 828	5 728 040	5 128	5 722 653	-259	0,00
COMLAW_YE	3 440 562	3 416 138	-24 424	3 447 069	6 507	3 442 087	1 525	0,04
COMLAWNO_DIV	1 749 631	1 741 004	-8 627	1 749 867	236	1 749 963	332	0,02

Tableau 5.2.1
Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

Caractéristique	Recensement	Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		
	chiffre	estimation	différence	estimation	différence	estimation	différence	écart (%)
COMLAWNO_OTHERS	3 793 163	3 778 466	-14 697	3 792 762	-401	3 793 576	413	0,01
COMLAWNO_SEP	688 126	685 664	-2 462	687 481	-645	688 056	-70	-0,01
COMLAWNO_SINGLE	13 624 243	13 453 334	-170 909	13 618 465	-5 778	13 622 342	-1 901	-0,01
COMLAWNO_SINGLE_GE15	7 944 684	7 824 151	-120 533	7 934 226	-10 458	7 942 819	-1 865	-0,02
COMLAWNO_WID	1 355 406	1 351 797	-3 609	1 355 414	8	1 355 557	151	0,01
COMLAWYE_MARRIED	16 577 048	16 508 308	-68 740	16 582 139	5 091	16 578 536	1 488	0,01
COUPLE	16 304 196	16 244 300	-59 896	16 317 065	12 869	16 313 204	9 008	0,06
EMPIN_SADA_GT50	10 115 337	10 071 665	-43 672	10 115 480	143	10 115 329	-8	0,00
EMPIN_SADA_LE50	10 116 607	10 006 750	-109 857	10 115 995	-612	10 116 600	-7	0,00
EMPIN_SADA_P0	13 762 510	13 661 692	-100 818	13 761 890	-620	13 762 525	15	0,00
EMPIN_SADA_P0_GE15	8 082 951	8 032 509	-50 442	8 077 651	-5 300	8 083 003	52	0,00
EMPIN_SADA_P25	5 058 711	5 008 873	-49 838	5 058 197	-514	5 059 199	488	0,01
EMPIN_SADA_P50	5 057 896	4 997 877	-60 019	5 057 797	-99	5 057 401	-495	-0,01
EMPIN_SADA_P75	5 057 676	5 030 940	-26 736	5 058 134	458	5 057 668	-8	0,00
EMPIN_SADA_P100	5 057 661	5 040 726	-16 935	5 057 347	-314	5 057 661	0	0,00
FEMALE	17 258 035	17 153 083	-104 952	17 262 580	4 545	17 258 007	-28	0,00
FEMALEGE15	14 489 305	14 407 839	-81 466	14 489 472	167	14 489 532	227	0,00
FEMALELT15	2 768 730	2 745 244	-23 486	2 773 108	4 378	2 768 475	-255	-0,01
HHADACSD (CSDxADA)	13 932 820	13 372 132	-560 688	13 932 820	0	13 932 820	0	0,00
HHINC_SADA_GT50	6 966 238	6 936 358	-29 880	6 966 557	319	6 966 487	249	0,00
HHINC_SADA_LE50	6 966 582	6 927 519	-39 064	6 966 263	-319	6 966 333	-249	0,00
HHINC_SADA_P25	3 483 439	3 458 080	-25 359	3 482 595	-844	3 482 799	-640	-0,02
HHINC_SADA_P50	3 483 143	3 469 439	-13 704	3 483 667	524	3 483 534	391	0,01

Tableau 5.2.1
Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

		Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		
Caractéristique	Recensement chiffre	estimation	différence	estimation	différence	estimation	différence	écart (%)
HHINC_SADA_P75	3 483 232	3 472 828	-10 404	3 483 857	625	3 483 764	532	0,02
HHINC_SADA_P100	3 483 006	3 463 530	-19 476	3 482 700	-306	3 482 723	-283	-0,01
HHSIZE1	3 939 604	3 932 270	-7 334	3 938 772	-832	3 937 579	-2 025	-0,05
HHSIZE2	4 801 532	4 788 730	-12 802	4 802 550	1 018	4 803 338	1 806	0,04
HHSIZE3	2 119 446	2 103 306	-16 140	2 119 554	108	2 120 058	612	0,03
HHSIZE4	1 927 188	1 915 750	-11 438	1 928 438	1 250	1 928 027	839	0,04
HHSIZE5	739 505	732 595	-6 910	742 931	3 426	739 636	131	0,02
HHSIZEGE5	1 145 050	1 123 820	-21 230	1 143 506	-1 544	1 143 819	-1 231	-0,11
HHSIZEGE6	405 545	391 225	-14 320	400 575	-4 970	404 183	-1 362	-0,34
INEFAM	28 720 122	28 545 611	-174 511	28 750 644	30 522	28 734 615	14 493	0,05
IR_LINK_NO	33 637 976	33 390 492	-247 484	33 637 564	-412	33 638 859	883	0,00
IR_LINK_YE	356 478	349 615	-6 863	355 801	-677	355 595	-883	-0,25
LIM_NO	29 187 451	28 997 962	-189 489	29 188 145	694	29 188 235	784	0,00
LIM_YE	4 807 003	4 742 145	-64 858	4 805 220	-1 783	4 806 219	-784	-0,02
LONEPAR	1 568 868	1 552 022	-16 846	1 563 594	-5 274	1 565 900	-2 968	-0,19
MALE	16 736 419	16 587 024	-149 395	16 730 785	-5 634	16 736 447	28	0,00
MALEGE15	13 825 590	13 703 085	-122 505	13 819 654	-5 936	13 825 399	-191	0,00
MALELT15	2 910 829	2 883 939	-26 890	2 911 130	301	2 911 048	219	0,01
MARRIED	13 136 486	13 092 169	-44 317	13 135 070	-1 416	13 136 449	-37	0,00
NOCLDFAM	3 998 054	3 993 087	-4 967	3 994 087	-3 967	3 999 849	1 795	0,04
NOINEFAM	5 274 332	5 194 497	-79 835	5 242 721	-31 611	5 259 839	-14 493	-0,27
NOINEFAMH-HSIZEGT1	1 334 728	1 262 226	-72 502	1 303 949	-30 779	1 322 260	-12 468	-0,93
NOTINFAM	6 038 982	5 945 010	-93 972	6 010 343	-28 639	6 030 190	-8 792	-0,15
NOTINFAMH-HSIZEGT1	2 099 378	2 012 740	-86 638	2 071 570	-27 808	2 092 611	-6 767	-0,32
OLN_BI	6 187 606	6 160 319	-27 287	6 187 142	-464	6 187 821	215	0,00
OLN_EN	23 157 389	22 959 992	-197 397	23 156 234	-1 155	23 157 275	-114	0,00
OLN_FR	4 015 036	3 989 439	-25 597	4 015 372	336	4 014 760	-276	-0,01
OLN_NO	634 423	630 357	-4 066	634 618	195	634 598	175	0,03
POBG2_1	14 784	14 945	161	14 798	14	14 841	57	0,38
POBG2_2	7 790	7 381	-409	7 390	-400	7 420	-370	-4,75
POBG2_3	27 825	27 490	-335	27 315	-510	27 564	-261	-0,94
POBG2_4	61 352	61 578	226	61 928	576	61 993	641	1,04
POBG2_5	6 443	6 217	-226	6 259	-184	6 238	-205	-3,18
POBG2_6	34 071	33 889	-182	33 847	-224	33 734	-337	-0,99
POBG2_7	68 871	68 511	-360	69 240	369	68 917	46	0,07
POBG2_8	4 069	4 012	-57	4 142	73	4 114	45	1,10

Tableau 5.2.1
Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

Caractéristique	Recensement	Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		
	chiffre	estimation	différence	estimation	différence	estimation	différence	écart (%)
POBG2_9	1 360	1 452	92	1 516	156	1 503	143	10,53
POBG2_10	32 812	32 871	59	32 625	-187	32 781	-31	-0,10
POBG2_11	100 655	100 325	-330	100 502	-153	100 519	-136	-0,14
POBG2_12	7 474	7 156	-318	7 298	-176	7 135	-339	-4,53
POBG2_13	2 230	2 034	-196	2 032	-198	2 039	-191	-8,56
POBG2_14	1 694	1 785	91	1 779	85	1 754	60	3,57
POBG2_15	1 405	1 508	103	1 505	100	1 491	86	6,10
POBG2_16	27 097	26 029	-1 068	26 732	-365	27 016	-81	-0,30
POBG2_17	22 097	22 521	424	22 907	810	22 708	611	2,77
POBG2_18	2 909	2 876	-33	2 896	-13	2 892	-17	-0,58
POBG2_19	22 252	22 054	-198	22 189	-63	22 073	-179	-0,80
POBG2_20	24 130	24 675	545	24 381	251	24 165	35	0,15
POBG2_21	873 285	868 958	-4 327	873 080	-205	873 254	-31	0,00
POBG2_22	109 015	108 270	-745	108 163	-852	108 570	-445	-0,41
POBG2_23	15 319	14 917	-402	15 094	-225	15 262	-57	-0,37
POBG2_24	42 218	42 067	-151	42 520	302	42 632	414	0,98
POBG2_25	14 794	13 853	-941	14 093	-701	14 366	-428	-2,90
POBG2_26	65 533	66 428	895	65 646	113	65 666	133	0,20
POBG2_27	56 296	56 892	596	56 841	545	56 950	654	1,16
POBG2_28	139 455	140 786	1 331	140 543	1 088	140 480	1 025	0,74
POBG2_29	70 138	69 530	-608	70 139	1	69 877	-261	-0,37
POBG2_30	33 952	34 553	601	33 707	-245	33 709	-243	-0,72
POBG2_31	103 283	102 569	-714	103 462	179	102 855	-428	-0,41
POBG2_32	186 698	186 575	-123	187 555	857	187 251	553	0,30
POBG2_33	112 041	110 677	-1 364	111 862	-179	111 726	-315	-0,28
POBG2_34	19 235	19 630	395	19 665	430	19 480	245	1,27
POBG2_35	72 931	72 295	-636	73 132	201	72 782	-149	-0,20
POBG2_36	74 350	73 931	-419	74 129	-221	74 570	220	0,30
POBG2_37	112 543	111 512	-1 031	112 225	-318	112 106	-437	-0,39
POBG2_38	86 917	87 053	136	87 116	199	87 155	238	0,27
POBG2_39	28 117	27 361	-756	27 542	-575	27 852	-265	-0,94
POBG2_40	124 170	123 731	-439	124 693	523	124 738	568	0,46
POBG2_41	2 234	2 204	-30	2 240	6	2 258	24	1,08
POBG2_42	154 606	153 966	-640	155 098	492	154 564	-42	-0,03
POBG2_43	609 741	607 631	-2 110	609 653	-88	609 822	81	0,01
POBG2_44	2 761	3 017	256	2 965	204	2 933	172	6,22
POBG2_45	60 546	60 919	373	60 721	175	60 639	93	0,15
POBG2_46	51 880	51 081	-799	51 305	-575	51 651	-229	-0,44
POBG2_47	125 745	124 733	-1 012	125 786	41	125 600	-145	-0,12
POBG2_48	348 474	348 730	256	348 747	273	348 490	16	0,00

Tableau 5.2.1
Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

		Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		
Caractéristique	Recensement chiffre	estimation	différence	estimation	différence	estimation	différence	écart (%)
POBG2_49	1 342	1 500	158	1 466	124	1 515	173	12,87
POBG2_50	1 044	1 064	20	1 068	24	1 078	34	3,23
POBG2_51	14 222	13 744	-478	13 803	-419	13 902	-320	-2,25
POBG2_53	266 119	265 374	-745	264 686	-1 433	265 339	-780	-0,29
POBG2_54	115 745	114 997	-748	115 501	-244	115 690	-55	-0,05
POBG2_55	20 529	20 181	-348	20 361	-168	20 472	-57	-0,28
POBG2_56	74 226	74 585	359	74 351	125	74 304	78	0,10
POBG2_57	27 373	28 272	899	28 058	685	27 813	440	1,61
POBG2_58	13 459	13 286	-173	13 275	-184	13 306	-153	-1,13
POBG2_59	40 929	40 823	-106	40 995	66	40 986	57	0,14
POBG2_60	123 382	125 358	1 976	123 900	518	123 828	446	0,36
POBG2_61	13 814	13 787	-27	13 814	0	13 816	2	0,02
POBG2_62	14 021	14 015	-6	14 103	82	13 920	-101	-0,72
POBG2_63	9 250	9 276	26	9 190	-60	8 962	-288	-3,11
POBG2_64	880 611	880 776	165	880 196	-415	880 351	-260	-0,03
POBG2_65	31 092	30 739	-353	30 820	-272	30 852	-240	-0,77
POBG2_66	292 617	294 336	1 719	292 849	232	292 496	-121	-0,04
POBG2_67	172 316	170 704	-1 612	172 312	-4	172 232	-84	-0,05
POBG2_68	21 518	21 146	-372	21 201	-317	21 317	-201	-0,94
POBG2_69	70 746	69 813	-933	70 611	-135	70 649	-97	-0,14
POBG2_70	38 949	37 477	-1 472	37 941	-1 008	38 369	-580	-1,49
POBG2_71	105 555	105 197	-358	105 258	-297	105 289	-266	-0,25
POBG2_M3	27 477 998	27 240 476	-237 522	27 478 602	604	27 479 831	1 833	0,01
POBG3_2	315 727	313 684	-2 043	315 604	-123	315 346	-381	-0,12
POBG3_3	52 168	51 502	-666	52 616	448	52 702	534	1,02
POBG3_4	144 752	144 355	-397	145 063	311	144 824	72	0,05
POBG3_5	172 246	173 386	1 140	171 963	-283	172 386	140	0,08
POBG3_6	1 026 914	1 021 550	-5 364	1 026 782	-132	1 027 359	445	0,04
POBG3_7	258 618	257 390	-1 228	258 524	-94	258 841	223	0,09
POBG3_8	526 099	524 544	-1 555	525 838	-261	525 741	-358	-0,07
POBG3_9	224 749	223 501	-1 248	225 241	492	224 445	-304	-0,14
POBG3_10	306 076	307 622	1 546	306 124	48	305 802	-274	-0,09
POBG3_12	48 365	47 675	-690	47 680	-685	48 041	-324	-0,67
POBG3_13	263 337	262 581	-756	261 909	-1 428	262 571	-766	-0,29
POBG3_14	288 406	288 209	-197	288 243	-163	288 093	-313	-0,11
POBG3_15	817 934	815 377	-2 557	818 608	674	818 128	194	0,02
POBG3_16	42 289	42 275	-14	42 511	222	42 489	200	0,47
POBG3_17	1 022 321	1 023 931	1 610	1 022 082	-239	1 022 232	-89	-0,01
POBG3_18	298 701	298 685	-16	298 557	-144	298 837	136	0,05
POBG3_19	208 492	209 073	581	209 002	510	208 606	114	0,05

Tableau 5.2.1
Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Canada

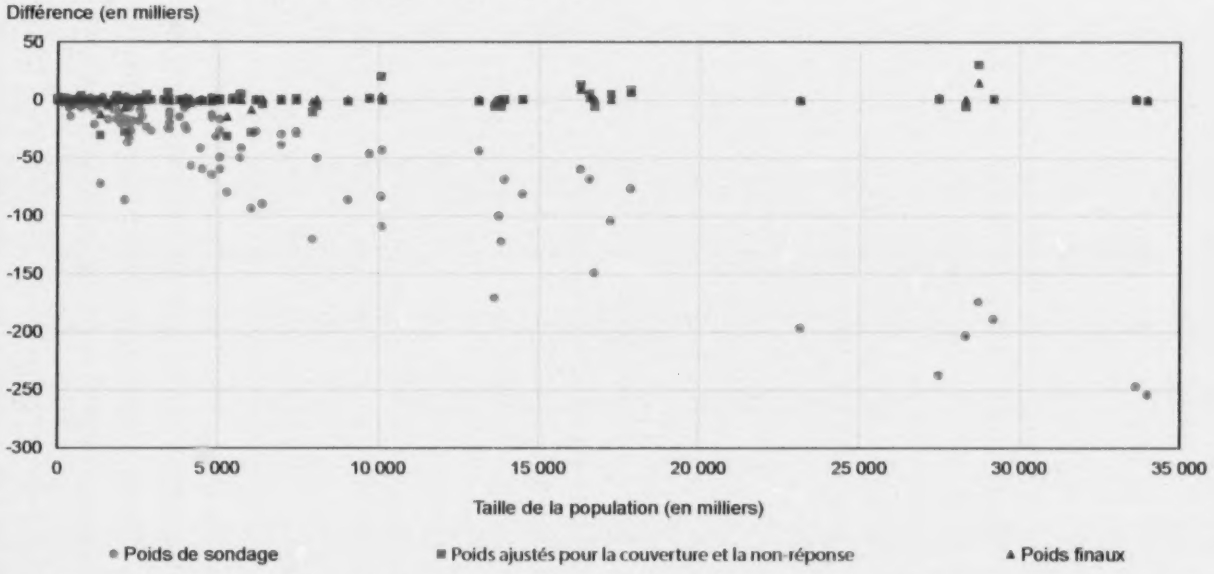
		Poids de sondage		Poids ajustés pour la couverture et la non-réponse		Poids finaux		
Caractéristique	Recensement chiffre	estimation	différence	estimation	différence	estimation	différence	écart (%)
POBG3_21	112 196	109 730	-2 466	111 021	-1 175	111 510	-686	-0,61
POBG3_22	212 520	211 823	-697	213 051	531	212 398	-122	-0,06
PPADACSD (CSDxADA)	33 994 454	32 767 748	-1 226 706	33 993 365	-1 089	33 994 454	0	0,00
SNGLDET	7 434 272	7 406 572	-27 700	7 435 570	1 298	7 434 328	56	0,00
TOTCFAM	9 720 966	9 674 172	-46 794	9 722 126	1 160	9 722 502	1 536	0,02
TOTHHLD	13 932 820	13 863 876	-68 944	13 932 820	0	13 932 820	0	0,00
TOTPERS	33 994 454	33 740 107	-254 347	33 993 365	-1 089	33 994 454	0	0,00
TPERGE15	28 314 895	28 110 924	-203 971	28 309 127	-5 768	28 314 932	37	0,00
YRIMD_1900	1 038 231	1 038 998	767	1 037 883	-348	1 038 201	-30	0,00
YRIMD_1981	273 907	275 157	1 250	274 631	724	274 167	260	0,09
YRIMD_1986	491 794	491 464	-330	491 411	-383	491 648	-146	-0,03
YRIMD_1991	726 588	725 927	-661	726 496	-92	726 876	288	0,04
YRIMD_1996	680 866	683 744	2 878	681 300	434	680 314	-552	-0,08
YRIMD_2001	849 532	846 726	-2 806	849 728	196	849 356	-176	-0,02
YRIMD_2006	961 099	956 822	-4 277	960 315	-784	960 960	-139	-0,01
YRIMD_2011	1 150 540	1 141 489	-9 051	1 150 015	-525	1 150 117	-423	-0,04
YRIMD_M5	343 899	339 304	-4 595	342 984	-915	342 983	-916	-0,27
YRIMG1_1981	765 701	766 620	919	766 042	341	765 815	114	0,01
YRIMG1_1991	1 407 454	1 409 672	2 218	1 407 796	342	1 407 190	-264	-0,02
YRIMG1_2001	1 810 631	1 803 548	-7 083	1 810 043	-588	1 810 317	-314	-0,02

Notes : Les estimations provenant des variables portant sur la connaissance des langues officielles, OLN_BI, OLN_EN et OLN_FR, ainsi que leur poids ont été calculés à l'aide des réponses existantes avant la mise à jour des données sur la langue. Une étude a démontré que l'utilisation de ces variables non corrigées comme contraintes de calage a un impact négligeable sur la qualité des estimations pondérées. Pour plus de renseignements sur cet impact, consultez l'annexe 1.8 du *Guide du Recensement de la population, 2016* (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/app-ann1-8-fra.cfm).

Tous les ménages dont le poids de sondage est égal à 1 sont exclus du processus de pondération. Ces ménages se trouvent dans une réserve indienne ou dans les UC de recensement par interview, ou correspondent à des ménages privés occupant un logement attenant à la structure d'un logement collectif.

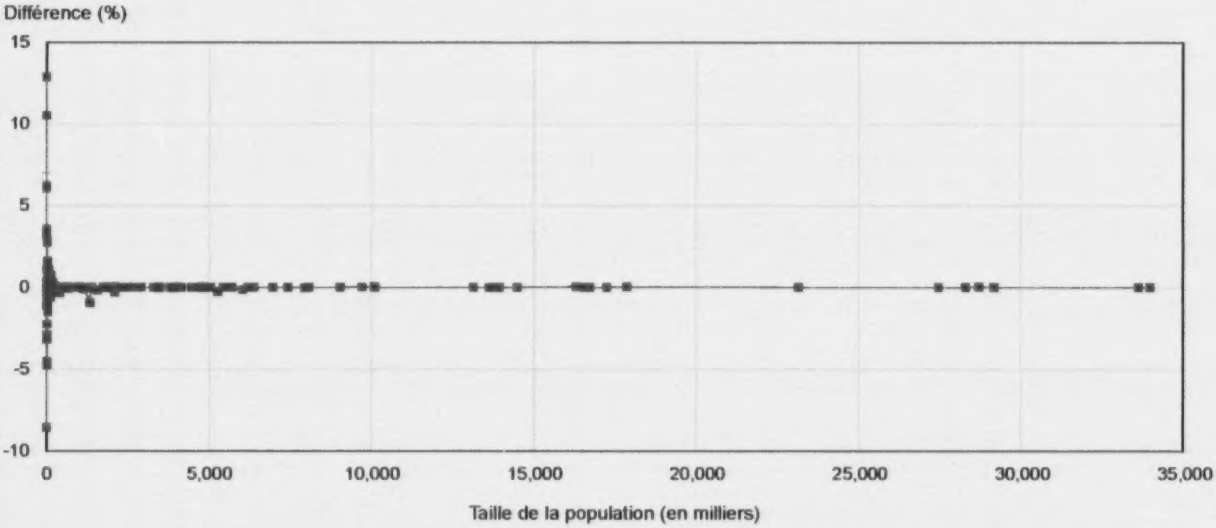
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

Graphique 5.2.2
Différences entre les chiffres du recensement et les estimations pondérées par le poids de sondage, ajustées pour la couverture et la non-réponse, et finales



Note : Tous les ménages ayant un poids de sondage de 1 ont été exclus du processus de pondération. Ces ménages se trouvaient dans une unité de collecte située dans une réserve indienne ou une unité de collecte de recensement par interview, ou étaient des ménages privés annexés à un logement collectif.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

Graphique 5.2.3
Écart entre les chiffres du recensement et les estimations finales, en tant que pourcentage des chiffres du recensement



Notes : Tous les ménages ayant un poids de sondage de 1 ont été exclus du processus de pondération. Ces ménages se trouvaient dans une unité de collecte située dans une réserve indienne ou une unité de collecte de recensement par interview, ou étaient des ménages privés annexés à un logement collectif.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

6. Estimation de la variance

L'erreur d'une estimation est la différence entre l'estimation et la valeur réelle de ce que l'on vise à estimer. Plusieurs sources d'erreur existent dans l'enquête-échantillon du questionnaire détaillé, dont l'erreur d'échantillonnage et l'erreur de non-réponse totale. L'erreur d'échantillonnage provient du fait que les estimations sont produites à partir d'observations provenant d'un échantillon et non du Recensement de la population. L'erreur de non-réponse totale survient quand des ménages sélectionnés dans l'échantillon ne répondent pas à l'enquête.

L'erreur a une composante aléatoire, mesurée par la variance, et une composante systématique, mesurée par le biais. La variance mesure la variabilité de l'estimation par rapport à l'estimation moyenne qui résulterait de répétitions hypothétiques du processus d'enquête. Cette variance peut être estimée à partir des données de l'échantillon. Le biais est la différence entre la valeur moyenne d'une estimation qui résulterait de répétitions hypothétiques du processus d'enquête et la valeur réelle de la caractéristique estimée. Les méthodes d'échantillonnage et d'estimation utilisées dans l'enquête-échantillon du questionnaire détaillé génèrent un biais négligeable.

Certaines méthodes d'estimation sont plus précises que d'autres pour estimer une caractéristique donnée de la population. Elles ont donc une incidence sur l'erreur. La variance estimée peut être utilisée pour produire plusieurs indicateurs de qualité couramment utilisés pour mesurer la précision d'une estimation. Par exemple, elle peut être utilisée pour calculer des erreurs-types, des intervalles de confiance et des coefficients de variation. L'erreur-type est l'indicateur de qualité produit pour les estimations provenant du questionnaire détaillé du Recensement de 2016. L'erreur-type correspond à la racine carrée de la variance.

Il est important de bien distinguer ces mesures de variabilité des autres mesures de la qualité qui ne sont pas des mesures de variabilité proprement dites. Des exemples de telles mesures sont les taux de réponse finaux présentés à la section 3.11 et le taux global de non-réponse au questionnaire détaillé du Recensement de 2016. En effet, le taux de réponse est un indicateur du risque associé à l'erreur de non-réponse des ménages. Le taux global de non-réponse est, quant à lui, un indicateur du risque d'erreur due à la non-réponse des ménages et à la non-réponse d'items aux questions individuelles. Veuillez consulter le chapitre 11 (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/chap11-fra.cfm) du *Guide du Recensement de la population, 2016* pour plus d'information (Statistique Canada, 2017).

Comme l'échantillon du questionnaire détaillé est stratifié géographiquement selon des strates à tirage partiel (les UC d'envoi par la poste et de listage-livraison) et des strates à tirage complet (les UC de recensement par interview), deux estimateurs de variance sont utilisés. Un premier estimateur de variance est utilisé pour estimer la variance dans les aires géographiques à tirage partiel (voir la section 6.3.1) et un deuxième estimateur est utilisé pour estimer la variance due à la non-réponse totale dans les aires à tirage complet (voir la section 6.3.2). Pour le reste du présent chapitre, le terme variance est employé pour désigner la variance d'échantillonnage et de non-réponse totale dans les aires géographiques à tirage partiel ou la variance de non-réponse totale dans les aires à tirage complet.

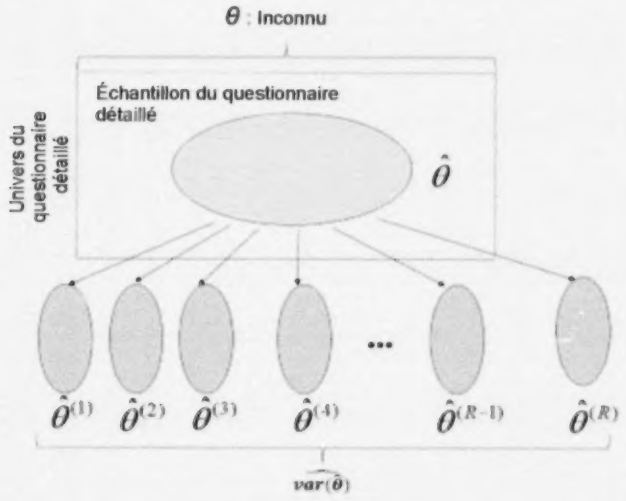
Les erreurs-types de diverses estimations et aires géographiques peuvent être consultées ou encore téléchargées à partir du Profil du recensement comprenant l'erreur-type, Canada, provinces et territoires, divisions de recensement (DR) et aires de diffusion agrégées (ADA), Recensement de 2016 (http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/adaprof/index.cfm?Lang=F&TABID=0). Les méthodes d'échantillonnage et d'estimation utilisées pour l'enquête au moyen du questionnaire détaillé de 2016 sont différentes de celles qui ont été utilisées pour l'Enquête nationale auprès des ménages de 2011. Comme l'ampleur et l'effet de la non-réponse étaient très différents en 2011, cela a une incidence sur la variabilité des estimations. Pour plus d'information, veuillez consulter la note méthodologique comparant les erreurs-types des estimations de 2016 à celles des estimations de 2011 et de 2006 (http://www12.statcan.gc.ca/census-recensement/2016/ref/se-et-fra.cfm) pour certaines variables.

6.1 Éléments à considérer pour choisir une méthode d'estimation de la variance

Étant donné le très grand nombre et la diversité d'estimations produites et afin de pouvoir établir des indicateurs de qualité pour ces mêmes estimations dans un délai raisonnable, un estimateur de variance par rééchantillonnage, dérivé de la méthode modifiée par répliques répétées partiellement équilibrées, a été utilisé (Judkins, 1990). La méthode consiste à tirer des échantillons (ou répliques) à partir de l'échantillon original. Des poids sont calculés pour chacune des répliques et ceux-ci subissent les mêmes ajustements de couverture, de non-réponse et de calage que l'échantillon original. Les poids résultants sont appelés poids de répliques. Des estimations sont ensuite produites pour chaque réplique, puis la variance est estimée à l'aide des estimations des répliques et de la moyenne de ces estimations.

La figure 6.1.1 donne un aperçu du concept d'estimation de la variance par répliques lorsque R échantillons sont utilisés.

Figure 6.1.1
Aperçu de l'estimation de la variance par répliques



La figure 6.1.1 donne un aperçu de la méthodologie d'estimation de la variance par répliques utilisée dans le cadre du Recensement de 2016. La méthode d'estimation de la variance par répliques simule la sélection de plusieurs échantillons afin d'estimer la variance d'échantillonnage.

Plus précisément, la figure montre l'univers du questionnaire détaillé représentant la population d'intérêt ainsi que l'échantillon du questionnaire détaillé. L'échantillon est situé à l'intérieur de l'univers pour indiquer qu'il correspond à un sous-ensemble de la population d'intérêt. Cet échantillon permet d'estimer une caractéristique de la population d'intérêt, par exemple « le nombre de personnes faisant partie d'une minorité visible ». Le symbole thêta est utilisé pour représenter la vraie valeur de cette caractéristique. Un accent circonflexe sur le thêta indique que la valeur représente une estimation de cette caractéristique. Cette valeur est appelée thêta chapeau.

Les R autres échantillons placés hors de l'univers sont reliés à l'échantillon du questionnaire détaillé par des flèches. Les flèches indiquent que ces échantillons sont tirés de l'échantillon du questionnaire détaillé. La caractéristique d'intérêt est réestimée à partir de ces R sous-échantillons. Les R valeurs de thêta chapeau, nommées thêta chapeau un, thêta chapeau deux, jusqu'à thêta chapeau R sont utilisées pour calculer la variance estimée de thêta chapeau.

Dans cette figure, nous définissons :

- thêta, θ, la vraie valeur de la caractéristique dans la population, qui peut être un total, une moyenne, un quantile, etc.;
- thêta chapeau, $\hat{\theta}$, la valeur de θ estimée à partir des poids principaux;
- thêta chapeau *r*, $\hat{\theta}^{(r)}$, la valeur de θ estimée à partir des poids de la réplique *r*, *r*=1, ..., R;
- thêta chapeau barre, $\bar{\hat{\theta}}$, la valeur moyenne des R estimations de répliques $\hat{\theta}^{(r)}$ et
- $\hat{\text{var}}(\hat{\theta})$, la valeur estimée de la variance de $\hat{\theta}$.

6.2 Estimateur de variance

L'estimateur par répliques choisi pour l'enquête-échantillon du questionnaire détaillé est dérivé de la méthode des demi-échantillons équilibrés de Fay (Judkins, 1990). La méthode détermine la création des répliques, le calcul des poids de répliques ainsi que le facteur multiplicatif utilisé pour estimer la variance.

Afin de produire les estimations de variance des estimations de l'échantillon du questionnaire détaillé, deux ensembles de poids de répliques ont été créés : un premier ensemble pour 32 poids de répliques et un deuxième ensemble pour 100 poids de répliques. L'ensemble de 32 poids de répliques a été produit pour estimer les erreurs-types de produits normalisés dont le calcul est sujet à des contraintes opérationnelles (c.-à-d. le besoin de diffuser un grand nombre d'erreurs-types dans un délai raisonnable), tandis que l'ensemble de 100 poids de répliques est mis à la disposition des analystes de Statistique Canada, des analystes des centres de données de recherche qui ont accès aux microdonnées ou des utilisateurs qui demandent des produits personnalisés, et ce, afin de leur fournir des estimateurs de variance plus précis.

L'estimateur de variance par répliques peut être calculé de deux façons, dont une plus conservatrice que l'autre. Il consiste à additionner les différences au carré entre les estimations de répliques, $\hat{\theta}^{(r)}$, et soit la moyenne des estimations des répliques, $\bar{\hat{\theta}}$, soit l'estimation provenant de l'échantillon principal, $\hat{\theta}$, et à multiplier cette somme par un certain facteur multiplicatif. La deuxième méthode, qui utilise l'estimation provenant de l'échantillon principal, est plus conservatrice. Dans le Système de spécification de produits assisté par ordinateur, utilisé pour la diffusion de statistiques, l'estimateur de variance est calculé à partir de la moyenne des estimations de répliques.

Par exemple, deux estimateurs de variance pour l'estimation d'un total calculé à partir d'un ensemble de R répliques sont donnés par les équations suivantes :

$$\hat{\text{Var}}_1\left(\hat{T}\right) = \frac{1}{R/2}\sum_{r=1}^{R}\left(\hat{T}^{(r)} - \bar{\hat{T}}\right)^2,$$

$$\hat{\text{Var}}_2\left(\hat{T}\right) = \frac{1}{R/2}\sum_{r=1}^{R}\left(\hat{T}^{(r)} - \hat{T}\right)^2$$

où

$$\hat{T} = \sum_{k\in s} w_k y_k,$$

$$\hat{T}^{(r)} = \sum_{k\in s} w_k^{(r)} y_k, \text{ et}$$

$$\bar{\hat{T}} = \sum_{r=1}^{R} \hat{T}^{(r)} / \mathrm{R}.$$

Le poids final de l'échantillon est représenté par w_k, $w_k^{(r)}$ est le poids final de la réplique r, y_k est la valeur de la caractéristique y pour l'unité k et s est l'échantillon du questionnaire détaillé.

Le nombre de degrés de liberté (DDL) de l'estimateur de variance est approximé par le nombre de différences au carré $\left(\hat{T}^{(r)} - \overline{\hat{T}}\right)^2$, soit 32 ou 100, de l'estimateur de variance. Le nombre de DDL donne un aperçu de la précision de l'estimateur de variance et il est utilisé lors du calcul d'intervalles de confiance pour les estimations provenant du questionnaire détaillé. Comme la méthode employée pour estimer la variance de l'estimation peut utiliser un petit nombre de répliques, on a recours à la loi de la probabilité de Student pour déterminer le quantile du niveau de confiance à utiliser lors du calcul de l'intervalle. Dans les cas où le nombre de ménages participant à l'estimation est plus grand que R, le nombre de DDL utilisé devrait être R. Dans les cas où le nombre de ménages est plus petit que R mais plus grand que ou égal à 20, le nombre de DDL à utiliser correspond au nombre de ménages. Si le nombre de ménages est plus petit que 20, alors il n'est pas recommandé de produire des intervalles de confiance.

6.3 Ajustement des poids de répliques

6.3.1 Unités de collecte d'envoi par la poste et de listage/livraison

Comme mentionné à la section 6.2, des poids de répliques sont calculés pour tous les ménages de l'échantillon du questionnaire détaillé. Les répliques sont partiellement équilibrées, c'est-à-dire que les répliques sont équilibrées par strate de rééchantillonnage, celles-ci étant créées en combinant des UC de façon à obtenir entre 600 et 1 800 ménages par strate de rééchantillonnage.

L'utilisation de la méthode modifiée des demi-échantillons équilibrés de Fay, décrite par Rao et Shao (1999), fait usage d'une valeur « epsilon » dans le calcul des poids de répliques afin de contrôler la perturbation des poids de répliques. Cette perturbation fait en sorte que tous les ménages échantillonnés participent à chacune des répliques, contrairement à d'autres méthodes d'estimation par répliques plus courantes. Cela facilite le calage des poids de répliques et, à l'occasion, le calcul des estimations ponctuelles sur chacune des répliques (p. ex. le dénominateur d'un estimateur de ratio pour une réplique donnée ne sera pas nul si le dénominateur correspondant n'était pas nul avec le poids final). L'ajout d'un facteur epsilon dans le calcul des poids de répliques a aussi permis de tenir compte de la grande fraction de sondage utilisée pour sélectionner l'échantillon du questionnaire détaillé. Les détails techniques de l'estimation de la variance sont donnés dans Devin et Verret (2016).

Les poids de répliques sont sujets aux mêmes ajustements que le poids de sondage de l'échantillon principal. Ils ont été ajustés pour la couverture et la non-réponse totale en suivant la même méthodologie que pour le poids de l'échantillon principal (voir la section 4.4). Ensuite, les poids de répliques résultants ont été calés aux chiffres du recensement, toujours en suivant la même méthodologie que pour le poids principal (voir la section 4.5).

6.3.2 Unités de collecte dans des réserves indiennes[8] et unités de collecte de recensement par interview

Comme le décrit le chapitre 2, tous les ménages situés dans des UC de réserves indiennes ou de recensement par interview ont été sélectionnés avec certitude. Par conséquent, ils avaient à l'origine un poids de sondage de 1. Aucun ajustement pour la couverture n'était nécessaire. Tous ces ménages ont été sélectionnés pour le questionnaire détaillé; il ne pouvait donc pas y avoir de différence de couverture entre le questionnaire abrégé et le questionnaire détaillé. La non-réponse totale dans ces aires a été traitée par le processus d'imputation de ménages entiers (IME), décrit au chapitre 3. En d'autres termes, les données d'un ménage non répondant ont été remplacées par les données d'un ménage répondant de la même UC (à l'exception des variables géographiques

8. L'exception à cette caractéristique était les unités situées dans des réserves indiennes et des établissements indiens partiellement dénombrés, qui ont été exclues de la population cible et dont le poids a été fixé à 0, sans autre modification à l'ensemble de données ou aux poids.

des non-répondants, que l'on connaissait déjà). La repondération n'a donc pas été nécessaire pour les ménages situés dans des UC de réserves indiennes ou de recensement par interview.

Le calage n'a pas non plus été nécessaire pour ces aires, car le questionnaire détaillé était un recensement. Par conséquent, tous les ménages situés dans des UC de réserves indiennes ou de recensement par interview ont conservé leur poids original de 1 dans le plan de pondération final. Pour plus de renseignements sur les réserves indiennes et les établissements indiens partiellement dénombrés, voir l'annexe 1.2 (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/app-ann1-2-fra.cfm) du *Guide du Recensement de la population, 2016*, n° 98-304-X au catalogue.

Bien qu'il n'y ait pas eu de variabilité d'échantillonnage dans les ménages situés dans des UC de réserves indiennes ou de recensement par interview, il y a eu de la variabilité liée à l'IME. L'estimation de la variance dans ces aires a été calculée selon une méthode semblable à celle pour le reste du pays, à quelques exceptions près. Premièrement, la probabilité de réponse selon la combinaison de la taille du ménage dans chaque division de recensement a été estimée en divisant le nombre de ménages répondants par le nombre de ménages visés par l'enquête. Ensuite, les poids de rééchantillonnage de base ont été créés comme pour le reste du pays, à l'exception du fait que tous les répondants dont la probabilité de réponse était égale à 1 ont été placés dans chaque réplique. Les répondants dont la probabilité de réponse estimée était inférieure à 1 n'ont pas été considérés comme des certitudes et ont été traités comme des éléments échantillonnés (c.-à-d. qu'ils ont été répartis aléatoirement entre les répliques). Les ménages non répondants imputés par l'IME ont également été répartis entre les répliques, et chacun s'est fait attribuer l'indicateur d'inclusion dans la réplique correspondant à son donneur d'une manière semblable à celle de Shao et Tang (2011). Ainsi, ce sont les poids plutôt que les valeurs qui ont varié d'une réplique à l'autre. Enfin, les poids de rééchantillonnage ont été calés en fonction du nombre de ménages et du nombre de personnes dans la SADA. Par conséquent, la variance estimée de ces deux quantités était égale à 0 au niveau de la SADA et aux niveaux plus agrégés, par exemple au niveau du Canada (puisque ces deux contraintes sont obligatoires dans le reste du pays).

7. Conclusion

Le retour du questionnaire détaillé obligatoire a apporté plusieurs caractéristiques souhaitables en matière d'échantillonnage et de pondération. En particulier, le plan d'échantillonnage a été grandement simplifié par rapport à celui de l'Enquête nationale auprès des ménages (ENM) de 2011 (voir Statistique Canada, 2015 pour de plus amples détails sur le plan d'échantillonnage de l'ENM de 2011). Une conséquence directe est la simplification de la procédure de pondération et une variation moins grande des poids finaux par rapport à celle des poids de l'ENM de 2011. En outre, la nature obligatoire du questionnaire détaillé de 2016 a sans doute été favorable au taux de réponse. Réduire le taux de non-réponse est toujours important, car cela peut avoir une incidence directe sur le biais des estimations et permettre de limiter autant que possible le biais de non-réponse.

La combinaison d'une fraction d'échantillonnage d'un sur quatre et du taux de réponse final global non pondéré de 96,1 % signifie que l'échantillon du questionnaire détaillé de 2016 a obtenu davantage de réponses dans l'ensemble que l'ENM de 2011 (qui avait une fraction d'échantillonnage d'un sur trois et un taux de réponse de 68,6 %). Les taux de réponse d'une aire géographique à l'autre étaient également plus homogènes. Le taux de réponse beaucoup plus élevé et plus uniforme a permis de simplifier les procédures de pondération.

Toutes ces caractéristiques, combinées avec une procédure de calage améliorée, ont fait en sorte que la différence maximale entre les chiffres du recensement et les estimations pour le poids final (14 493) correspondait environ au dixième de celle pour l'ENM de 2011 (166 801). L'amélioration de la procédure de calage était attribuable à la sélection d'un moins grand nombre de variables de calage, à des aires de pondération plus grandes (c.-à-d. les SADA) et à une meilleure sélection de contraintes (p. ex. la priorité accordée aux contraintes qui partagent la population en deux parts à peu près égales).

Ces améliorations ont permis de réduire l'étendue des poids finaux (par rapport à l'ENM de 2011), ce qui a fait diminuer l'erreur-type pour la plupart des caractéristiques. Pour les variables liées aux contraintes de calage qui étaient sélectionnées souvent, la variance a été réduite par la méthode de calage. Pour les variables faiblement liées aux contraintes de calage, la variance a néanmoins été limitée par la faible étendue des poids.

Enfin, la diffusion des poids de rééchantillonnage semble indiquer que les utilisateurs peuvent facilement estimer l'erreur-type pour toute variable pertinente. Lors de l'ENM de 2011, la méthode d'estimation de la variance sélectionnée — la linéarisation par série de Taylor, combinée avec un facteur de correction vers le haut (tiré de la simulation de Monte Carlo) — avait rendu difficile le calcul de l'erreur-type pour les produits non normalisés.

Annexe A – Glossaire

Voici les définitions des termes, variables et concepts principaux du recensement mentionnés dans le présent document. Les utilisateurs peuvent également consulter le *Dictionnaire, Recensement de la population, 2016* (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/index-fra.cfm), nº 98-301-X au catalogue pour plus de renseignements.

Aire de diffusion (AD) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo021-fra.cfm) : Une aire de diffusion (AD) est une petite unité géographique relativement stable formée d'un ou de plusieurs îlots de diffusion avoisinants dont la population moyenne est de 400 à 700 habitants d'après les données du Programme du Recensement de la population précédent. Il s'agit de la plus petite région géographique normalisée pour laquelle toutes les données du recensement sont diffusées. Les AD couvrent tout le territoire du Canada.

Aire de diffusion agrégée (ADA) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo053-fra.cfm) : L'aire de diffusion agrégée (ADA) est une nouvelle région géographique de diffusion créée pour le Recensement de 2016. Les ADA couvrent l'ensemble du pays et, dans la mesure du possible, ont une population de 5 000 à 15 000 habitants selon les chiffres de population du recensement précédent. Les ADA sont créées à partir de régions géographiques de diffusion existantes et sont constituées à partir de secteurs de recensement (SR), de subdivisions de recensement (SDR) ou d'aires de diffusion (AD). Les ADA respectent les limites provinciales et territoriales ainsi que les limites des divisions de recensement (DR), des régions métropolitaines de recensement (RMR) et des agglomérations de recensement (AR) avec secteurs de recensement (SR).

L'objectif de l'ADA consiste à garantir la disponibilité de données du recensement, dans la mesure du possible, pour toutes les régions du Canada.

Division de recensement (DR) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo008-fra.cfm) : Groupe de municipalités voisines les unes des autres qui sont réunies pour des besoins de planification régionale et de gestion de services communs (comme les services de police et d'ambulance). Ces groupes sont créés selon les lois en vigueur dans certaines provinces du Canada. Division de recensement (DR) est le terme général de régions créées en vertu des lois provinciales (comme comté, municipalité régionale de comté et *regional district*) ou des régions équivalentes. Dans d'autres provinces et dans les territoires, où les lois ne prévoient pas de telles régions, Statistique Canada définit des régions équivalentes à des fins statistiques en collaboration avec ces provinces et territoires. Les divisions de recensement sont des régions géographiques intermédiaires entre la municipalité (subdivision de recensement) et la province/territoire.

Famille de recensement (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/fam004-fra.cfm) : On définit la « Famille de recensement » comme suit : un couple marié et les enfants, le cas échéant, du couple et/ou de l'un ou l'autre des conjoints; un couple en union libre et les enfants, le cas échéant, du couple et/ou de l'un ou l'autre des partenaires; ou un parent seul, peu importe son état matrimonial, habitant avec au moins un enfant dans le même logement et cet ou ces enfants. Tous les membres d'une famille de recensement particulière habitent le même logement. Un couple peut être de sexe opposé ou de même sexe. Les enfants peuvent être des enfants naturels, par le mariage, par l'union libre ou par adoption, peu importe leur âge ou leur état matrimonial, du moment qu'ils habitent dans le logement sans leur propre conjoint marié, partenaire en union libre ou enfant. Les petits-enfants habitant avec leurs grands-parents, alors qu'aucun des parents n'est présent, constituent également une famille de recensement.

Famille économique (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/fam011-fra.cfm) : La « famille économique » désigne un groupe de deux personnes ou plus habitant dans le même logement et apparentées par le sang, le mariage, l'union libre, l'adoption ou une relation de famille d'accueil. Un couple peut être de sexe opposé ou de même sexe. Par définition, toutes les personnes qui sont membres d'une famille de recensement sont également membres d'une famille économique. Voici des exemples du vaste concept de famille économique : deux familles de recensement apparentées et vivant ensemble sont considérées comme une famille économique; les frères et soeurs habitant ensemble qui ne sont pas membres d'une famille de recensement sont considérés comme une famille économique; et les nièces ou les neveux habitant avec leur tante ou leur oncle sont considérés comme une famille économique.

Logement (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/dwelling-logements004-fra.cfm) : Un logement est défini comme un ensemble de pièces d'habitation. Dans le cadre du recensement, on reconnaît deux genres de logements, soit les logements collectifs et les logements privés. Les premiers sont des logements de nature institutionnelle, communautaire ou commerciale. Les autres désignent un ensemble distinct de pièces d'habitation, auquel on a accès par une entrée privée soit à partir de l'extérieur de l'immeuble, soit à partir d'un hall, d'un foyer, d'un vestibule ou d'un escalier commun situé à l'intérieur de l'immeuble. Il faut qu'on puisse emprunter l'entrée menant au logement sans passer par les pièces d'habitation d'une autre personne ou d'un autre groupe de personnes.

Logement collectif (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/dwelling-logements002-fra.cfm) : Établissement commercial, institutionnel ou communautaire que l'agent recenseur peut identifier comme tel grâce à une enseigne ou en s'adressant à la personne qui en a la charge, à un résident, à un voisin, etc. Sont inclus dans cette catégorie les pensions et maisons de chambres, les hôtels, motels et établissements pour touristes, les établissements de soins infirmiers, les hôpitaux, les résidences de personnel, les bases militaires, les camps de travailleurs, les prisons, les foyers collectifs, etc.

Logement privé (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/dwelling-logements005-fra.cfm) : « Logement privé » s'entend d'un ensemble séparé de pièces d'habitation possédant une entrée privée soit à partir de l'extérieur de l'immeuble, soit à partir d'un hall, d'un foyer, d'un vestibule ou d'un escalier commun situé à l'intérieur de l'immeuble. Il faut qu'on puisse emprunter l'entrée menant au logement sans passer par les pièces d'habitation d'une autre personne ou d'un autre groupe de personnes.

Le logement doit répondre aux deux conditions qui le rendent propre à l'habitation durant toute l'année :

1. avoir une source de chauffage ou d'énergie (ce dont témoignent les cheminées, les lignes électriques, les conduites ou les compteurs d'huile ou de gaz, les génératrices, les cordes de bois, les ampoules électriques, les thermopompes ou les panneaux solaires)
2. fournir un espace clos permettant de s'abriter des intempéries comme en témoignent la présence de murs d'enceinte et d'un toit ainsi que de portes et fenêtres qui protègent contre le vent, la pluie et la neige.

Les logements qui ne répondent pas aux conditions les rendant propres à l'habitation durant toute l'année sont des logements marginaux. Les logements privés sont classés comme logements privés ordinaires et logements marginaux occupés. Les logements privés ordinaires se subdivisent à leur tour en trois grandes catégories : les logements occupés (par des résidents habituels), les logements occupés uniquement par des résidents étrangers et/ou par des personnes présentes temporairement et les logements inoccupés. Les logements marginaux sont classés comme logements occupés par des résidents habituels ou comme logements occupés par des résidents étrangers et/ou par des personnes présentes temporairement. Les logements marginaux inoccupés le 10 mai 2016 ne font pas partie du parc immobilier.

Logement privé occupé par des résidents habituels (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/dwelling-logements006-fra.cfm) : Logement privé occupé de façon permanente par une personne ou un groupe de personnes. Sont également inclus dans cette catégorie les logements privés dont les résidents habituels sont temporairement absents le 10 mai 2016. Sauf indication contraire, toutes les données présentées dans les produits sur le logement ont trait aux logements privés occupés par des résidents habituels, et non aux logements privés inoccupés ou aux logements occupés uniquement par des résidents étrangers et/ou par des personnes présentes temporairement.

Ménage (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/households-menage007-fra.cfm) : Ménage réfère à une personne ou à un groupe de personnes qui occupent le même logement et n'ont pas de domicile habituel ailleurs au Canada ou à l'étranger. Le logement peut être un logement collectif ou un logement privé. Le ménage peut se composer d'un groupe familial (famille de recensement), de deux familles ou plus partageant le même logement, d'un groupe de personnes non apparentées ou d'une personne vivant seule. Les membres du ménage qui sont temporairement absents le jour de référence sont considérés comme faisant partie de leur ménage habituel.

Ménage privé (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/households-menage014-fra.cfm) : « Ménage privé » s'applique à une personne ou un groupe de personnes qui occupent le même logement et qui n'ont pas de domicile habituel ailleurs au Canada ou à l'étranger. Les ménages sont divisés en deux sous-univers selon qu'ils occupent un logement collectif (ménage collectif) ou un logement privé (ménage privé).

Dans le cadre du recensement, les ménages sont classés en trois catégories : les ménages privés, les ménages collectifs et les ménages à l'extérieur du Canada.

À moins d'indications contraires, toutes les données contenues dans les produits du recensement se rapportent aux ménages privés seulement.

Secteur de recensement (SR) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo013-fra.cfm) : Les secteurs de recensement (SR) sont de petites régions géographiques relativement stables qui comptent habituellement une population de moins de 10 000 habitants d'après les données du Programme du Recensement de la population précédent. Ils sont situés dans les régions métropolitaines de recensement et dans les agglomérations de recensement dont la population du noyau urbain était égale ou supérieure à 50 000 habitants lors du recensement précédent.

Un comité de spécialistes locaux (par exemple, des planificateurs, des travailleurs sociaux, des travailleurs du secteur de la santé et des éducateurs) délimite initialement les secteurs de recensement de concert avec Statistique Canada. Une fois qu'une région métropolitaine de recensement (RMR) ou qu'une agglomération de recensement (AR) a été subdivisée en secteurs de recensement, les secteurs de recensement sont maintenus même si, ultérieurement, la population du noyau de la RMR ou de l'AR devient inférieure à 50 000 habitants.

Subdivision de recensement (SDR) (http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo012-fra.cfm) : La subdivision de recensement (SDR) est un terme générique qui désigne les municipalités (telles que définies par les lois provinciales ou territoriales) ou les territoires considérés comme étant des équivalents municipaux à des fins statistiques (p. ex. les réserves indiennes, les établissements indiens et les territoires non organisés). Le statut de municipalité est défini par les lois en vigueur dans chaque province et territoire du Canada.

Unité de collecte (UC)[9] : Les unités de collecte sont de petites unités géographiques utilisées pour la collecte des données du recensement. Tout le territoire du Canada est couvert par les unités de collecte.

Super aire de diffusion agrégée (SADA)[10] : Les super aires de diffusion agrégées ont été créées spécifiquement pour la pondération du Recensement de 2016 en respectant des règles préétablies, certaines obligatoires et d'autres optionnelles. Dans tous les cas, les SADA ont été créées en regroupant des ADA et elles respectent les limites provinciales et territoriales. Toutes les personnes vivant dans des UC de recensement par interview sont exclues de la population des SADA. Dans la mesure du possible, les SADA ont été construites afin d'avoir une population de 50 000 à 150 000 personnes. Les divisions de recensement (DR) qui ont une population de 40 000 à 50 000 personnes forment leur propre SADA. De plus, les SADA respectent, si possible, en ordre de priorité, les limites des divisions de recensement (DR), des régions métropolitaines de recensement (RMR), des agglomérations de recensement (AR) et des subdivisions de recensement (SDR). Enfin, les SADA devaient être formées par le regroupement d'ADA adjacentes lorsque cela était faisable, et être aussi compactes que possible.

9. Cette définition n'est pas dans le *Dictionnaire, Recensement de la population, 2016*, car le dictionnaire est composé principalement de termes liés à la diffusion tandis qu'« unité de collecte » est un terme relatif à la collecte.

10. Cette définition n'est pas dans le *Dictionnaire, Recensement de la population, 2016*, car le dictionnaire est composé principalement de termes liés à la diffusion tandis qu'« unité de collecte » est un terme relatif à la collecte.

Annexe B – Historique de l'échantillonnage dans le cadre du recensement canadien

L'échantillonnage a été utilisé pour la première fois lors du recensement canadien de 1941. Un questionnaire sur le logement a été rempli à tous les 10 logements. Les réponses aux 27 questions de ce questionnaire ont été intégrées aux données de la section sur les personnes et le ménage du questionnaire sur la population correspondant à ces logements. Cela a permis de faire le recoupement des caractéristiques de l'échantillon et des caractéristiques de base. Lors de ce même recensement, on a utilisé l'échantillonnage à l'étape du dépouillement pour obtenir des estimations provisoires des gains des salariés, de la répartition de la population en âge de travailler et de la composition des familles au Canada. Dans ce cas, l'échantillon était constitué du dixième de tous les secteurs de dénombrement du Canada et tous les questionnaires sur la population de ces secteurs ont été dépouillés à l'avance.

L'échantillonnage a également été utilisé pour le Recensement du logement de 1951. Cette fois, on a choisi chaque cinquième logement (ceux dont le numéro d'identification se terminait par 2 ou par 7) et le questionnaire sur le logement comportait 24 questions. Au cours du Recensement de 1961, on a prélevé un échantillon de 20 % des ménages privés et on a demandé aux personnes âgées de 15 ans et plus de ces ménages de remplir un questionnaire-échantillon sur la population comportant des questions sur la migration interne, la fécondité et le revenu. L'échantillonnage n'a pas été utilisé au cours des recensements de moindre envergure de 1956 et de 1966.

Plusieurs innovations importantes ont été apportées aux méthodes de dénombrement lors du Recensement de 1971. La première a été l'utilisation de l'autodénombrement pour la majorité de la population au lieu de la méthode traditionnelle de recensement par interview. Ce changement découlait des résultats de plusieurs études effectuées au Canada et ailleurs (Fellegi, 1964; Hansen et autres, 1959), selon lesquels la présence d'un recenseur avait un effet important sur la variance des chiffres du recensement lors d'un recensement par interview. On prévoyait donc que l'autodénombrement permettrait de réduire la variance des chiffres du recensement attribuable à la présence du recenseur et, en même temps, donnerait au répondant le temps voulu pour remplir le questionnaire en toute confidentialité; par conséquent, on s'attendait à ce que les réponses soient plus exactes que par le passé.

Un second changement survenu au Recensement de 1971 et qui le différencie de tous les recensements précédents a été la modification du contenu. Le nombre de sujets traités et de questions posées était plus élevé qu'aux recensements antérieurs. Une étude des coûts, du fardeau du répondant et de la rapidité d'exécution par rapport à la qualité des données dans le cadre d'un recensement par autodénombrement et par échantillonnage a motivé la décision de recueillir toutes les données, sauf celles portant sur certaines caractéristiques de base, auprès d'un échantillon d'un tiers de la population au Recensement de 1971. Sauf dans les régions les plus éloignées du Canada, le tiers des ménages privés a reçu le « questionnaire complet » comportant toutes les questions du recensement, tandis que le reste des ménages privés a reçu le « questionnaire abrégé » qui ne contenait que les questions de base sur le nom, le lien avec le chef de ménage, le sexe, la date de naissance, l'état matrimonial, la langue maternelle, le type de logement, le mode d'occupation, le nombre de pièces, l'alimentation en eau, les installations sanitaires ainsi que certaines questions en rapport avec la couverture du recensement. Tous les ménages vivant dans des régions éloignées prédésignées et dans des logements collectifs ont reçu le questionnaire complet. Le rapport statistique Sampling in the Census (Bureau fédéral de la statistique, 1968) donne une description plus détaillée des projets d'application de la méthode d'échantillonnage au Recensement de 1971.

Le Recensement de 1976 comportait nettement moins de contenu que celui de 1971. En outre, le questionnaire de 1976 ne contenait pas les questions qui posent le plus de difficulté de collecte (p. ex. les questions sur le revenu) ni celles pour lesquelles les frais de codage sont le plus élevés (p. ex. les questions sur la profession, l'industrie et le lieu de travail). Par conséquent, les avantages de l'échantillonnage en matière de coûts et d'allégement du fardeau du répondant étaient moins évidents qu'au Recensement de 1971. Néanmoins, après avoir évalué les économies éventuelles liées à diverses fractions d'échantillonnage et étudié les conséquences sur le plan des relations publiques du retour au dénombrement intégral (l'échantillonnage ayant fait ses preuves en 1971), Statistique Canada a décidé d'appliquer en 1976 la même méthode d'échantillonnage qu'en 1971.

Au Recensement de 1981, on a utilisé la plupart des mêmes méthodes qu'en 1971 et 1976, sauf que le taux d'échantillonnage est passé du tiers au cinquième des ménages privés occupés. Les études réalisées à ce moment-là ont révélé que la diminution de la qualité des données (mesurée au moyen de la variance) qui en résultait serait acceptable et qu'elle ne serait pas suffisamment importante pour annuler les avantages découlant de la réduction des coûts, du fardeau du répondant et des délais d'exécution (Royce, 1983). Le taux d'échantillonnage du cinquième des ménages a été la méthode retenue pour tous les recensements de 1981 à 2006.

En 2011, l'information qui était auparavant fournie au moyen du questionnaire détaillé du recensement (à participation obligatoire) a été recueillie dans le cadre de l'Enquête nationale auprès des ménages (ENM), à participation volontaire. En raison de ce changement, chaque ménage a dû répondre aux 10 questions contenues dans le questionnaire du Recensement de 2011, tandis que 30 % des ménages ont été sélectionnés pour répondre à l'ENM. Par ailleurs, un ménage sur trois n'ayant pas répondu à l'ENM a été sous-échantillonné aux fins de suivi. La fraction d'échantillonnage accrue de 30 % a été appliquée en prévision d'un taux de réponse plus faible à l'ENM. Pour le Recensement de 2016, le gouvernement a de nouveau rendu obligatoire le questionnaire détaillé du recensement pour remplacer l'ENM. La fraction d'échantillonnage a été augmentée à un sur quatre en 2016, par rapport à un sur cinq pour le questionnaire détaillé du recensement de 2006, en vue de réduire le risque que le taux de réponse ne revienne pas à ses niveaux élevés du passé.

Annexe C – Contraintes utilisées ou exclues dans le processus de pondération[11]

Voici la liste des contraintes potentielles définies au niveau des ADA et au niveau des SADA. Il y a 270 contraintes potentielles au niveau des ADA et 200 contraintes potentielles au niveau des SADA. Les contraintes qui ne sont pas définies dans une ADA ou une SADA ne sont pas comptabilisées dans le tableau. Pour cette raison, on ne retrouve pas 4 283 ou 4 180 ADA et 408 SADA comme dans les tableaux précédents.

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
ADULTCF	Adultes faisant partie d'une famille de recensement	SADA	105	302	SADA	138	270
AGE00_14	Personnes âgées de 0 à 14 ans	SADA	288	120	LES DEUX	4 473	115
AGE4	Personnes âgées de 0 à 4 ans	SADA	70	338	LES DEUX	868	3 720
AGE9	Personnes âgées de 5 à 9 ans	SADA	64	344	LES DEUX	836	3 752
AGE14	Personnes âgées de 10 à 14 ans	SADA	52	356	LES DEUX	820	3 768
AGE15_24	Personnes âgées de 15 à 24 ans	SADA	310	98	SADA	390	18
AGE15_29	Personnes âgées de 15 à 29 ans	...	...	...	ADA	4 084	96
AGE19	Personnes âgées de 15 à 19 ans	SADA	268	140	LES DEUX	822	3 766
AGE24	Personnes âgées de 20 à 24 ans	SADA	252	156	LES DEUX	885	3 703
AGE25_34	Personnes âgées de 25 à 34 ans	SADA	373	35	SADA	401	7
AGE29	Personnes âgées de 25 à 29 ans	SADA	324	84	LES DEUX	1 406	3 182
AGE30_49	Personnes âgées de 30 à 49 ans	...	...	...	ADA	4 148	32
AGE34	Personnes âgées de 30 à 34 ans	SADA	348	60	LES DEUX	1 575	3 013
AGE35_44	Personnes âgées de 35 à 44 ans	SADA	360	48	SADA	408	0
AGE39	Personnes âgées de 35 à 39 ans	SADA	369	39	LES DEUX	1 455	3 133
AGE44	Personnes âgées de 40 à 44 ans	SADA	389	19	LES DEUX	1 384	3 204

11. Voir l'annexe D au sujet des contraintes applicables à Wood Buffalo.

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
AGE45_54	Personnes âgées de 45 à 54 ans	SADA	406	2	SADA	408	0
AGE49	Personnes âgées de 45 à 49 ans	SADA	403	5	LES DEUX	1 656	2 932
AGE50_64	Personnes âgées de 50 à 64 ans	...	...	...	ADA	4 026	154
AGE54	Personnes âgées de 50 à 54 ans	SADA	405	3	LES DEUX	1 918	2 670
AGE55_64	Personnes âgées de 55 à 64 ans	SADA	390	18	SADA	402	6
AGE59	Personnes âgées de 55 à 59 ans	SADA	385	23	LES DEUX	1 683	2 905
AGE64	Personnes âgées de 60 à 64 ans	SADA	371	37	LES DEUX	1 491	3 097
AGE65PL	Personnes âgées de 65 ans et plus	SADA	390	18	LES DEUX	4 361	227
AGE74	Personnes âgées de 65 à 74 ans	SADA	400	8	LES DEUX	2 128	2 460
AGE75PL	Personnes âgées de 75 ans et plus	SADA	391	17	LES DEUX	2 109	2 479
APTLT5	Ménages vivant en appartement dans un immeuble de moins de cinq étages	SADA	406	2	LES DEUX	2 669	1 919
CHILD	Enfants faisant partie d'une famille de recensement	SADA	7	400	SADA	39	369
CHILDFAM	Familles de recensement avec enfants	SADA	16	391	LES DEUX	3 145	1 443
COMLAW_YE	Personnes vivant en union libre	SADA	252	155	LES DEUX	2 708	1 880
COMLAWNO_ DIV	Personnes divorcées ne vivant pas en union libre	SADA	405	2	LES DEUX	1 181	3 407
COMLAWNO_ OTHERS	Personnes ne vivant pas en union libre qui sont soient divorcées, séparées ou veuves	SADA	399	8	LES DEUX	2 954	1 634

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
COMLAWNO_SEP	Personnes séparées et ne vivant pas en union libre	SADA	399	8	LES DEUX	609	3 979
COMLAWNO_SINGLE	Personnes célibataires (jamais mariées) et ne vivant pas en union libre	SADA	204	203	LES DEUX	3 496	1 092
COMLAWNO_SINGLE_LT15	Personnes célibataires âgées de moins de 15 ans (jamais mariées) et ne vivant pas en union libre	...	...	...	LES DEUX	4 473	115
COMLAWNO_WID	Personnes veuves ne vivant pas en union libre	SADA	398	9	LES DEUX	763	3 825
COMLAWYE_MARRIED	Personnes mariées ou vivant en union libre	SADA	200	207	LES DEUX	3 241	1 347
COUPLE	Personnes faisant partie d'un couple (mariage, partenaire)	SADA	111	296	SADA	139	269
EMPIN_GT50	Personnes dont le revenu d'emploi annuel est au-dessus du 50e percentile de l'ADA	...	...	...	ADA	4 110	70
EMPIN_LE50	Personnes dont le revenu d'emploi annuel est égal ou inférieur au 50e percentile de l'ADA	...	...	...	ADA	4 094	86
EMPIN_P0	Personnes dont le revenu d'emploi annuel est nul, au niveau des ADA	...	...	...	ADA	4 129	51
EMPIN_P0_GE15	Personnes âgées de 15 ans et plus dont le revenu d'emploi annuel est nul, au niveau des ADA	...	...	...	ADA	4 069	111

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
EMPIN_P0_LT15	Personnes âgées de moins de 15 ans dont le revenu d'emploi annuel est nul, au niveau des ADA	...	...	...	ADA	4 083	97
EMPIN_P25	Personnes dont le revenu d'emploi annuel est non nul et inférieur ou égal au 25e percentile de l'ADA	...	...	...	ADA	2 636	1 544
EMPIN_P50	Personnes dont le revenu d'emploi annuel est au-dessus du 25e percentile et inférieur ou égal au 50e percentile de l'ADA	...	...	...	ADA	2 619	1 561
EMPIN_P75	Personnes dont le revenu d'emploi annuel est au-dessus du 50e percentile et inférieur ou égal au 75e percentile de l'ADA	...	...	...	ADA	3 713	467
EMPIN_P100	Personnes dont le revenu d'emploi annuel est au-dessus du 75e percentile de l'ADA	...	...	...	ADA	3 718	462
EMPIN_SADA_GT50	Personnes dont le revenu d'emploi annuel est au-dessus du 50e percentile de la SADA	SADA	405	3	SADA	407	1
EMPIN_SADA_LE50	Personnes dont le revenu d'emploi annuel est égal ou inférieur au 50e percentile de la SADA	SADA	406	2	SADA	407	1

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
EMPIN_SADA_P0	Personnes dont le revenu d'emploi annuel est nul, au niveau des SADA	SADA	407	1	SADA	406	2
EMPIN_SADA_P0_GE15	Personnes âgées de 15 ans et plus dont le revenu d'emploi annuel est nul, au niveau des SADA	...	...	...	SADA	392	16
EMPIN_SADA_P0_LT15	Personnes âgées de moins de 15 ans dont le revenu d'emploi annuel est nul, au niveau des SADA	...	...	...	SADA	390	18
EMPIN_SADA_P25	Personnes dont le revenu d'emploi annuel est non nul et inférieur ou égal au 25e percentile de la SADA	SADA	341	67	SADA	192	216
EMPIN_SADA_P50	Personnes dont le revenu d'emploi annuel est au-dessus du 25e percentile et inférieur ou égal au 50e percentile de la SADA	SADA	339	69	SADA	192	216
EMPIN_SADA_P75	Personnes dont le revenu d'emploi annuel est au-dessus du 50e percentile et inférieur ou égal au 75e percentile de la SADA	SADA	386	22	SADA	407	1
EMPIN_SADA_P100	Personnes dont le revenu d'emploi annuel est au-dessus du 75e percentile de la SADA	SADA	387	21	SADA	408	0
FEMALE	Femmes	SADA	39	369	LES DEUX	4 400	188

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
FEMALEGE15	Femmes âgées de 15 ans et plus	SADA	386	22	LES DEUX	3 054	1 534
FEMALELT15	Femmes âgées de moins de 15 ans	SADA	35	373	LES DEUX	2 942	1 646
HHADA (ADA)	Ménages dans l'ADA	ADA	4 043	195	...	...	...
HHADACSD	Ménages qui chevauchent la subdivision de recensement et l'ADA	...	...	...	ADA	4 938	1 994
HHINC_GT50	Ménages dont le revenu annuel est au-dessus du 50e percentile de l'ADA	...	...	...	ADA	4 138	42
HHINC_LE50	Ménages dont le revenu annuel est inférieur ou égal au 50e percentile de l'ADA	...	...	...	ADA	4 138	42
HHINC_P25	Ménages dont le revenu annuel est inférieur ou égal au 25e percentile de l'ADA	...	...	...	ADA	4 022	158
HHINC_P50	Ménages dont le revenu annuel est au-dessus du 25e percentile et inférieur ou égal au 50e percentile de l'ADA	...	...	...	ADA	4 024	156
HHINC_P75	Ménages dont le revenu annuel est au-dessus du 50e percentile et inférieur ou égal au 75e percentile de l'ADA	...	...	...	ADA	4 014	166
HHINC_P100	Ménages dont le revenu annuel est au-dessus du 75e percentile de l'ADA	...	...	...	ADA	4 014	166

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
HHINC_SADA_GT50	Ménages dont le revenu annuel est au-dessus du 50e percentile de la SADA	SADA	407	0	SADA	407	1
HHINC_SADA_LE50	Ménages dont le revenu annuel est inférieur ou égal au 50e percentile de la SADA	SADA	407	0	SADA	407	1
HHINC_SADA_P25	Ménages dont le revenu annuel est inférieur ou égal au 25e percentile de la SADA	SADA	407	0	SADA	407	1
HHINC_SADA_P50	Ménages dont le revenu annuel est au-dessus du 25e percentile et inférieur ou égal au 50e percentile de la SADA	SADA	407	0	SADA	407	1
HHINC_SADA_P75	Ménages dont le revenu annuel est au-dessus du 50e percentile et inférieur ou égal au 75e percentile de la SADA	SADA	402	5	SADA	407	1
HHINC_SADA_P100	Ménages dont le revenu annuel est au-dessus du 75e percentile de la SADA	SADA	402	5	SADA	407	1
HHSIZE1	Ménages de 1 personne	SADA	28	379	LES DEUX	766	3 822
HHSIZE2	Ménages de 2 personnes	SADA	114	293	LES DEUX	3 985	603
HHSIZE3	Ménages de 3 personnes	SADA	394	13	LES DEUX	3 634	954
HHSIZE4	Ménages de 4 personnes	SADA	376	31	LES DEUX	3 639	949
HHSIZE5	Ménages de 5 personnes	SADA	131	276	LES DEUX	991	3 597

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
HHSIZEGE5	Ménages de 5 personnes et plus	SADA	15	392	LES DEUX	386	4 202
HHSIZEGE6	Ménages de 6 personnes et plus	SADA	13	394	LES DEUX	277	4 311
INEFAM	Personnes dans une famille économique	SADA	51	356	SADA	111	297
IR_LINK_NO	Personnes qui n'ont pas pu être appariées au Registre des Indiens	SADA	207	200	LES DEUX	515	4 073
IR_LINK_YE	Personnes qui ont pu être appariées au Registre des Indiens	SADA	207	200	LES DEUX	451	4 137
LIM_NO	Personnes ne faisant pas partie d'une famille économique à faible revenu	SADA	399	8	LES DEUX	3 267	1 321
LIM_YE	Personnes faisant partie d'une famille économique à faible revenu	SADA	399	8	LES DEUX	3 267	1 321
LONEPAR	Parents seuls	SADA	118	289	SADA	149	259
MALE	Hommes	SADA	39	369	LES DEUX	4 400	188
MALEGE15	Hommes âgés de 15 ans et plus	SADA	295	113	LES DEUX	3 054	1 534
MALELT15	Hommes âgés de 14 ans et moins	SADA	53	355	LES DEUX	2 938	1 650
MARRIED	Personnes mariées	SADA	346	62	LES DEUX	2 761	1 827
NOCLDFAM	Familles de recensement sans enfants	SADA	40	367	LES DEUX	1 774	2 814
NOINEFAM	Personnes dans un ménage ne faisant pas partie d'une famille économique	SADA	51	356	SADA	111	297

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
NOINEFAMH-HSIZEEQ1	Personnes provenant d'un ménage de taille 1 ne faisant pas partie d'une famille économique	...	...	...	SADA	103	305
NOINEFAMH-HSIZEGT1	Personnes provenant d'un ménage de plus d'une personne ne faisant pas partie d'une famille économique	...	...	...	SADA	160	248
NOTINFAM	Personnes ne faisant pas partie d'une famille de recensement	SADA	75	332	SADA	143	265
NOTINFAMH-HSIZEGT1	Personne dans un ménage de plus d'une personne ne faisant pas partie d'une famille de recensement	...	...	...	SADA	98	310
OLN_BI	Langues officielles – Anglais et Français	SADA	374	33	LES DEUX	1 874	2 714
OLN_EN	Langue officielle – Anglais	SADA	78	329	LES DEUX	559	4 029
OLN_FR	Langue officielle – Français	SADA	74	333	LES DEUX	782	3 806
OLN_NO	Langue officielle – Aucune	SADA	224	183	LES DEUX	455	4 133
POBG2_1	Pays de naissance - Albanie	SADA	10	299	LES DEUX	64	4 524
POBG2_2	Pays de naissance - Andorre, Espagne et Gibraltar	SADA	22	366	LES DEUX	101	4 487
POBG2_3	Pays de naissance - Australie et Nouvelle-Zélande	SADA	13	383	LES DEUX	191	4 397

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_4	Pays de naissance - Autriche, Allemagne et Liechtenstein	SADA	25	382	LES DEUX	262	4 326
POBG2_5	Pays de naissance - Estonie, Lettonie et Lituanie	SADA	17	332	LES DEUX	62	4 526
POBG2_6	Pays de naissance - Belgique et Pays-Bas	SADA	14	393	LES DEUX	241	4 347
POBG2_7	Pays de naissance - Belize, El Salvador, Guatemala et Honduras	SADA	26	381	LES DEUX	173	4 415
POBG2_8	Pays de naissance - Bhoutan	SADA	3	122	LES DEUX	10	4 578
POBG2_9	Pays de naissance - Botswana, Lesotho et Swaziland	SADA	13	254	LES DEUX	11	4 577
POBG2_10	Pays de naissance - Brésil	SADA	15	389	LES DEUX	138	4 450
POBG2_11	Pays de naissance - Bulgarie et Romanie	SADA	45	361	LES DEUX	179	4 409
POBG2_12	Pays de naissance - Birmanie (Myanmar)	SADA	6	285	LES DEUX	39	4 549
POBG2_13	Pays de naissance - Canada	SADA	34	335	LES DEUX	40	4 548

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_14	Pays de naissance - Anguilla, Bermudes, Îles Caïmans, Montserrat, Îles Turks-et-Caïcos et Îles Vierges britanniques	SADA	9	303	LES DEUX	26	4 562
POBG2_15	Pays de naissance - Guadeloupe, Martinique et Saint-Martin (partie française)	SADA	4	181	LES DEUX	17	4 571
POBG2_16	Pays de naissance - République démocratique du Congo et République du Congo	SADA	16	366	LES DEUX	85	4 503
POBG2_17	Pays de naissance - Cameroun, Gabon, République Centrafricaine et Tchad	SADA	16	350	LES DEUX	71	4 517
POBG2_18	Pays de naissance - Angola et Sao Tomé-et-Principe	SADA	14	289	LES DEUX	44	4 544
POBG2_19	Pays de naissance - Kazakhstan, Kirghizistan, Tadjikistan, Turkménistan et Ouzbékistan	SADA	17	356	LES DEUX	89	4 499
POBG2_20	Pays de naissance - Chili	SADA	19	384	LES DEUX	162	4 426
POBG2_21	Pays de naissance - Chine, Hong Kong, Macao et Taïwan	SADA	65	342	LES DEUX	269	4 319

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_22	Pays de naissance - Colombie, Équateur et Pérou	SADA	69	338	LES DEUX	256	4 332
POBG2_23	Pays de naissance - Costa Rica, Nicaragua et Panama	SADA	16	379	LES DEUX	127	4 461
POBG2_24	Pays de naissance - République Tchèque, Hongrie et Slovaquie	SADA	20	380	LES DEUX	161	4 427
POBG2_25	Pays de naissance - Burundi et Rwanda	SADA	14	340	LES DEUX	68	4 520
POBG2_26	Pays de naissance - Érythrée, Kenya, Ouganda, Tanzanie et Zambie	SADA	33	358	LES DEUX	159	4 429
POBG2_27	Pays de naissance - Comores, Djibouti, Madagascar, Malawi, Maurice, Seychelles, Somalie et Zimbabwe	SADA	28	378	LES DEUX	150	4 438
POBG2_28	Pays de naissance - Bélarus, Moldova, Fédération de Russie et Ukraine	SADA	43	364	LES DEUX	176	4 412
POBG2_29	Pays de naissance - Égypte, Soudan et Soudan du Sud	SADA	22	377	LES DEUX	131	4 457
POBG2_30	Pays de naissance - Éthiopie	SADA	28	325	LES DEUX	111	4 477
POBG2_31	Pays de naissance - France, Luxembourg et Monaco	SADA	27	380	LES DEUX	197	4 391

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_32	Pays de naissance - Cambodge, Laos et Viet Nam	SADA	59	348	LES DEUX	149	4 439
POBG2_33	Pays de naissance - Cuba, République Dominicaine et Haïti	SADA	34	372	LES DEUX	185	4 403
POBG2_34	Pays de naissance - Grèce	SADA	16	374	LES DEUX	131	4 457
POBG2_35	Pays de naissance - Guyana et Suriname	SADA	45	327	LES DEUX	140	4 448
POBG2_36	Pays de naissance - Italie, Saint-Marin et Saint-Siège (État de la Cité du Vatican)	SADA	54	349	LES DEUX	241	4 347
POBG2_37	Pays de naissance - Bahamas, Jamaïque et Porto Rico	SADA	76	314	LES DEUX	221	4 367
POBG2_38	Pays de naissance - Antigua-et-Barbuda, Aruba, Barbade, Bonaire, Saint-Eustache et Saba, Curaçao, Dominique, Grenade, Saint-Kitts-et-Nevis, Sainte-Lucie, Saint-Martin (partie néerlandaise), Saint-Vincent-et-les Grenadines, Trinité-et-Tobago et Îles Vierges des États-Unis	SADA	39	357	LES DEUX	186	4 402
POBG2_39	Pays de naissance - Japon	SADA	27	369	LES DEUX	175	4 413

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_40	Pays de naissance - Corée du Nord et Corée du Sud	SADA	24	376	LES DEUX	120	4 468
POBG2_41	Pays de naissance - Libéria	SADA	12	274	LES DEUX	27	4 561
POBG2_42	Pays de naissance - Algérie, Libye, Maroc et Tunisie	SADA	23	382	LES DEUX	134	4 454
POBG2_43	Pays de naissance - Brunéi Darussalam, Indonésie, Malaisie, Philippines, Singapour et Timor-Leste	SADA	87	320	LES DEUX	266	4 322
POBG2_44	Pays de naissance - Malte	SADA	10	270	LES DEUX	36	4 552
POBG2_45	Pays de naissance - Mexique	SADA	17	390	LES DEUX	183	4 405
POBG2_46	Pays de naissance - Arabie saoudite, Bahreïn, Émirats arabes unis, Qatar et Yémen	SADA	24	356	LES DEUX	156	4 432
POBG2_47	Pays de naissance - Liban, Syrie	SADA	33	373	LES DEUX	164	4 424
POBG2_48	Pays de naissance - Afghanistan, Cisjordanie et bande de Gaza (Palestine), Chypre, Iran, Iraq, Israël, Jordanie, Koweït, Oman et Turquie	SADA	94	309	LES DEUX	218	4 370
POBG2_49	Pays de naissance - Mongolie	SADA	3	161	LES DEUX	12	4 576

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_50	Pays de naissance - Mozambique	SADA	18	243	LES DEUX	20	4 568
POBG2_51	Pays de naissance - Népal	SADA	11	315	LES DEUX	33	4 555
POBG2_54	Pays de naissance - Pologne	SADA	64	339	LES DEUX	222	4 366
POBG2_55	Pays de naissance - Région de l'Océanie (exclut Australie et Nouvelle-Zélande)	SADA	13	348	LES DEUX	57	4 531
POBG2_56	Pays de naissance - Portugal	SADA	52	343	LES DEUX	222	4 366
POBG2_57	Pays de naissance - Argentine, Bolivie, Paraguay et Uruguay	SADA	19	383	LES DEUX	154	4 434
POBG2_58	Pays de naissance - Îles Åland, Danemark, Îles Féroé, Finlande, Islande, Norvège et Suède	SADA	14	383	LES DEUX	158	4 430
POBG2_59	Pays de naissance - Namibie et République d'Afrique du Sud	SADA	13	373	LES DEUX	146	4 442
POBG2_60	Pays de naissance - Sri Lanka	SADA	27	342	LES DEUX	93	4 495
POBG2_61	Pays de naissance - Suisse	SADA	18	387	LES DEUX	165	4 423
POBG2_62	Pays de naissance - Thaïlande	SADA	20	379	LES DEUX	138	4 450
POBG2_63	Pays de naissance - Arménie, Azerbaïdjan et Géorgie	SADA	9	324	LES DEUX	48	4 540

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG2_64	Pays de naissance - Bangladesh, Inde et Pakistan	SADA	38	364	LES DEUX	161	4 427
POBG2_65	Pays de naissance - Union des républiques socialistes soviétiques, ancienne	SADA	20	365	LES DEUX	138	4 450
POBG2_66	Pays de naissance - Guernesey, Irlande, Jersey, Île de Man, Royaume-Uni et Sercq	SADA	10	397	LES DEUX	127	4 461
POBG2_67	Pays de naissance - États-Unis	SADA	220	187	LES DEUX	443	4 145
POBG2_68	Pays de naissance - Venezuela	SADA	16	364	LES DEUX	111	4 477
POBG2_69	Pays de naissance - Gambie, Ghana, Nigéria et Sierra Leone	SADA	13	371	LES DEUX	106	4 482
POBG2_70	Pays de naissance - Bénin, Burkina Faso, Côte d'Ivoire, Guinée, Mali, Mauritanie, Niger, Sénégal et Togo	SADA	9	372	LES DEUX	73	4 515
POBG2_71	Pays de naissance - Bosnie-Herzégovine, Croatie, Kosovo, République de Macédoine, Monténégro, Serbie et Slovénie	SADA	55	346	LES DEUX	221	4 367
POBG3_1	Pays de naissance - Canada	SADA	34	335	LES DEUX	40	4 548

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG3_2	Pays de naissance - Antilles et Bermudes	SADA	154	253	LES DEUX	374	4 214
POBG3_3	Pays de naissance - Afrique centrale	SADA	27	375	LES DEUX	142	4 446
POBG3_4	Pays de naissance - Amérique centrale	SADA	128	279	LES DEUX	305	4 283
POBG3_5	Pays de naissance - Afrique de l'Est	SADA	120	287	LES DEUX	300	4 288
POBG3_6	Pays de naissance - Asie de l'Est	SADA	205	202	LES DEUX	739	3 849
POBG3_7	Pays de naissance - Europe de l'Est (Union des républiques socialistes soviétiques, ancienne)	SADA	180	227	LES DEUX	381	4 207
POBG3_8	Pays de naissance - Asie de l'Ouest et centrale et Moyen-Orient	SADA	191	215	LES DEUX	470	4 118
POBG3_9	Pays de naissance - Afrique du Nord	SADA	89	318	LES DEUX	262	4 326
POBG3_10	Pays de naissance - Europe du Nord	SADA	247	160	LES DEUX	461	4 127
POBG3_12	Pays de naissance - Océanie	SADA	47	356	LES DEUX	215	4 373
POBG3_14	Pays de naissance - Amérique du Sud	SADA	176	231	LES DEUX	364	4 224
POBG3_15	Pays de naissance - Asie du Sud-Est	SADA	249	158	LES DEUX	652	3 936
POBG3_16	Pays de naissance - Afrique du Sud	SADA	17	370	LES DEUX	158	4 430

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
POBG3_17	Pays de naissance - Asie du Sud	SADA	218	185	LES DEUX	647	3 941
POBG3_18	Pays de naissance - Europe du Sud	SADA	182	225	LES DEUX	407	4 181
POBG3_19	Pays de naissance - Union des républiques socialistes soviétiques, ancienne	SADA	112	295	LES DEUX	307	4 281
POBG3_20	Pays de naissance – États-Unis	SADA	220	187	LES DEUX	443	4 145
POBG3_21	Pays de naissance - Afrique de l'Ouest	SADA	68	339	LES DEUX	211	4 377
POBG3_22	Pays de naissance - Europe de l'Ouest	SADA	195	212	LES DEUX	449	4 139
PPADA (ADA)	Personnes dans l'ADA	ADA	42	4 196	...	...	...
PPADACSD	Personnes dans le chevauchement géographique de la subdivision de recensement et de l'ADA	...	...	...	ADA	4 881	2 051
SNGLDET	Ménages vivant dans un logement individuel non attenant	SADA	392	16	LES DEUX	3 704	884
TOTCFAM	Familles de recensement	SADA	333	74	LES DEUX	1 672	2 916
TOTHHLD	Ménages	SADA	408	0	LES DEUX	4 588	0
TOTPERS	Personnes	SADA	408	0	LES DEUX	4 588	0
TPERGE15	Personnes âgées de 15 ans et plus	SADA	288	120	LES DEUX	4 474	114
YRIMD_1900	Personnes dont l'année d'immigration est avant 1981	SADA	360	47	SADA	376	32

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
YRIMD_1981	Personnes dont l'année d'immigration est entre 1981 et 1985	SADA	229	178	SADA	277	131
YRIMD_1986	Personnes dont l'année d'immigration est entre 1986 et 1990	SADA	224	183	SADA	276	132
YRIMD_1991	Personnes dont l'année d'immigration est entre 1991 et 1995	SADA	238	169	SADA	293	115
YRIMD_1996	Personnes dont l'année d'immigration est entre 1996 et 2000	SADA	231	176	SADA	290	118
YRIMD_2001	Personnes dont l'année d'immigration est entre 2001 et 2005	SADA	259	148	SADA	309	99
YRIMD_2006	Personnes dont l'année d'immigration est entre 2006 et 2010	SADA	258	149	SADA	310	98
YRIMD_2011	Personnes dont l'année d'immigration est entre 2011 et 2016	SADA	300	107	SADA	333	75
YRIMD_M5	Personnes dont l'année d'immigration est manquante	...	...	...	SADA	239	169
YRIMG1_1900	Personnes dont l'année d'immigration est avant 1981	SADA	360	47	SADA	376	32
YRIMG1_1981	Personnes dont l'année d'immigration est entre 1981 et 1990	SADA	310	97	SADA	338	70
YRIMG1_1991	Personnes dont l'année d'immigration est entre 1991 et 2000	SADA	320	87	SADA	350	58

Tableau C.1
Statistiques sur l'utilisation des contraintes de calage par contrainte

		Ajustement pour la couverture et la non-réponse			Calage final		
Variable/ Contrainte	Description	Région	Nombre de contraintes calées	Nombre de contraintes exclues	Région	Nombre de contraintes calées	Nombre de contraintes exclues
YRIMG1_2001	Personnes dont l'année d'immigration est entre 2001 et 2010	SADA	346	61	SADA	369	39
YRIMG1_2011	Personnes dont l'année d'immigration est entre 2011 et 2016	SADA	300	107	SADA	333	75

... n'ayant pas lieu de figurer
SADA = Super aire de diffusion agrégée.
ADA = Aire de diffusion agrégée.
LES DEUX = SADA/ADA.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

Annexe D – Wood Buffalo

Introduction

Le 1er mai 2016, un incendie de forêt s'est déclaré dans le sud-ouest de Fort McMurray, en Alberta. Le 3 mai, il a ravagé la région, détruisant bon nombre de maisons et d'immeubles et causant la plus importante évacuation de résidents en raison des feux de forêt de l'histoire de l'Alberta. Statistique Canada a alors décidé de suspendre la collecte des données du recensement (appelée la « collecte des données sur le terrain ») dans les zones évacuées. Plusieurs réponses avaient alors déjà été reçues des zones évacuées.

Statistique Canada a appliqué un ensemble de mesures pour veiller à ce que les résidents de la subdivision de recensement (SDR) de Wood Buffalo (appelée *specialized municipality of Wood Buffalo* ou Wood Buffalo) soient inclus dans le Recensement de la population de 2016. Les données des zones évacuées ont été tirées d'une combinaison de sources : la collecte directe de données sur le terrain et les sources de données administratives. Statistique Canada a travaillé en étroite collaboration avec les autorités provinciales et locales en Alberta afin d'obtenir un accès aux dossiers administratifs pour aider à valider les données dérivées des sources de données administratives disponibles à Statistique Canada.

La présente annexe fournit des détails supplémentaires sur la méthodologie utilisée dans la SDR de Wood Buffalo à la suite de l'incendie. Elle couvre en particulier les procédures de pondération et d'estimation. De plus amples détails sur les procédures de collecte se trouvent dans l'annexe 1.4 (http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/app-ann1-4-fra.cfm) du *Guide du Recensement de la population, 2016*, nº 98-304-X au catalogue.

Date de référence

En ce qui concerne le Recensement de 2016, la date de référence pour la déclaration des données était le 10 mai 2016. Pour les résidents des zones évacuées lors de l'incendie, la date de référence était le 1er mai 2016. On considère que cette nouvelle date représentait mieux la situation dans la SDR de Wood Buffalo. Par conséquent, un léger écart pouvait exister dans les réponses reçues entre le début de l'incendie et la reprise de la collecte des données sur le terrain en août. On considère ce problème comme mineur.

Données administratives

Dans la mesure du possible et lorsqu'aucune réponse directe n'avait été reçue pour un logement donné, des données provenant de diverses sources de données administratives ont été utilisées selon une date de référence se rapprochant le plus possible de mai 2016, pour les variables comme le nom, la date de naissance, le sexe et l'état matrimonial.

Comme les fichiers de données administratives n'avaient pas le même contenu en matière de langues que celui recueilli dans le questionnaire du recensement, des couplages de données ont été effectués entre les sources administratives et la base de données du Recensement de 2011. En ce qui concerne les couplages réussis, les réponses de 2011 aux questions sur la langue ont servi de substituts pour les questions sur la langue du Recensement de 2016. Les réponses aux questions du recensement pour lesquelles aucun renseignement comparable ne pouvait être tiré des fichiers de données administratives, comme le lien avec la Personne 1 et l'union libre, ont été dérivées au cours du traitement des données.

Collecte des données

En août 2016, la collecte des données a repris à Wood Buffalo, et des agents recenseurs ont fait du porte-à-porte pour remplir des questionnaires du recensement. Les efforts étaient axés sur la collecte des données pour les logements faisant partie de l'échantillon du questionnaire détaillé (un logement sur quatre). Pour améliorer davantage la qualité des données, la collecte des données sur le terrain a été effectuée pour les logements dans les secteurs pour lesquels aucune donnée administrative n'était disponible ainsi que pour les logements collectifs. Dans les secteurs où les agents recenseurs avaient préparé une liste des logements et livré des documents du recensement (secteurs de listage/livraison), les données étaient recueillies sur le terrain pour tous les logements.

Traitement

Si une réponse au recensement avait été obtenue des résidents d'un logement, les données de ces répondants avaient la préséance sur toutes données administratives disponibles. Pour les cas restants, pendant le traitement des données et aux fins du calcul des taux de réponse, les données provenant de sources administratives étaient considérées comme des réponses dans la même mesure qu'une réponse directe obtenue par des méthodes de collecte traditionnelles. Les procédures décrites au chapitre 3 ont été appliquées aux données reçues directement des répondants comme aux données provenant de sources administratives.

Qualité des données

Pour les chiffres de population et des logements, les données de la SDR de Wood Buffalo ont été soumises aux mêmes évaluations de la qualité que l'ensemble des données du recensement. Statistique Canada a réalisé une activité supplémentaire préalable à la validation après avoir combiné les données recueillies sur le terrain et les données de sources administratives. Cette étape supplémentaire a été effectuée pour attester que les méthodes de rechange conçues pour cette situation exceptionnelle permettaient d'obtenir des résultats satisfaisants.

Estimation

Pour éviter que l'ajustement de calage n'entraîne un biais dans les estimations, les données de l'échantillon et les données du recensement devaient être comparables sur le plan conceptuel. À Wood Buffalo, le taux de réponse de collecte pour le questionnaire abrégé (2A) était plus faible que celui pour le questionnaire détaillé (2A-L). Cet écart s'explique par l'objectif qui consistait, dans la mesure du possible, à remplir les questionnaires 2A à l'aide de sources de données administratives et à remplir les questionnaires 2A-L à l'aide de données de collecte.

Dans le cas du questionnaire 2A, 27 % des données provenaient de réponses directes, 67 % provenaient de sources de données administratives et 6 % provenaient de l'Enquête sur la classification des logements (ECL) [voir le chapitre 2]. Dans le cas du questionnaire 2A-L, 79 % des données provenaient de réponses directes, 8 % provenaient de sources administratives du questionnaire 2A, 6 % provenaient d'une combinaison de sources administratives du questionnaire 2A et de questionnaires 2A-L imputés, 6 % provenaient de questionnaires imputés 2A au moyen de l'ECL, et 1 %, uniquement de réponses au questionnaire 2A.

Certains concepts du recensement sont plus difficiles à obtenir de sources de données administratives, notamment ceux qui ont trait à la famille, à l'état matrimonial, à l'union libre et au revenu du ménage. Par conséquent, la différence de la contribution des sources de données administratives au questionnaire 2A et au questionnaire 2A-L a entraîné, comme on s'y attendait, des différences importantes entre les chiffres du recensement et les estimations de l'échantillon (avant l'ajustement des poids).

Lorsque des vérifications de la pondération ont été effectuées, on a constaté que la distribution de la pondération était atypique ou bimodale pour Wood Buffalo. On attribue cette situation au fait qu'il y avait davantage de ménages de taille 1 que prévu dans la base de données administratives (une base de données établie par Statistique Canada à l'aide de sources administratives utilisées pour remplir le questionnaire 2A) par rapport aux données provenant du questionnaire détaillé du recensement.

Des écarts considérables ont également été constatés entre les caractéristiques tirées du Registre des Indiens et les données administratives sur l'immigration couplées aux données du recensement. Dans ces cas, les différences étaient en partie dues au fait que ces données peuvent être couplées beaucoup plus facilement à la base de données administratives qu'aux questionnaires retournés, le Registre des Indiens et les données administratives sur l'immigration ayant servi à l'établissement de la base de données administratives.

Pour résoudre ces problèmes, certaines contraintes ont été désactivées pour Wood Buffalo afin d'éviter d'introduire un biais dans les estimations. Les contraintes désactivées portent sur les caractéristiques relatives à la famille, à l'état matrimonial, à l'union libre, à la taille du ménage et au revenu du ménage, et sur les caractéristiques tirées du Registre des Indiens et des données administratives sur l'immigration couplées aux dossiers du recensement.

Le tableau suivant montre les différences entre les chiffres du recensement et les estimations établies selon le questionnaire détaillé pour les caractéristiques qui présentent de grands écarts. Il est à noter que les chiffres du recensement provenaient largement de sources administratives. C'est particulièrement le cas en ce qui a trait aux variables liées au questionnaire 2A-L, comme l'année d'immigration, où les « chiffres du recensement » étaient en fait strictement des données administratives.

Tableau D.1
Différences et écarts entre les chiffres du recensement et les estimations du questionnaire détaillé, Wood Buffalo

Variable/contrainte	Description	Chiffres du recensement (provenant surtout de sources administratives)	Estimations (2A-L)	Différence relative (%)	Différence
ADULTCF	Adultes faisant partie d'une famille de recensement	32 505	35 215	-8,34	-2 710
COMLAW_YE	Personnes vivant en union libre	5 986	7 754	-29,53	-1 768
COMLAWNO_SEP	Personnes séparées et ne vivant pas en union libre	1 678	1 080	35,63	598
HHINC_P25	Ménages dont le revenu annuel est inférieur ou égal au 25e percentile de l'ADA	6 341	5 466	13,79	875
LIM_YE	Personnes faisant partie d'une famille économique à faible revenu	3 134	2 462	21,44	672
LONEPAR	Parents seuls	2 547	2 113	17,05	434
NOCLDFAM	Familles de recensement sans enfants	6 167	7 034	-14,05	-867
NOINEFAM	Personnes dans un ménage ne faisant pas partie d'une famille économique	13 714	10 202	25,61	3 512
NOTINFAM	Personnes ne faisant pas partie d'une famille de recensement	17 342	12 875	25,76	4 467
HHSIZE1	Ménages de 1 personne	5 879	4 646	20,98	1 233
HHSIZEGE6	Ménages de 6 personnes et plus	1 450	1 096	24,40	354
IR_LINK_YE	Personnes qui ont pu être appariées au Registre des Indiens	2 261	1 853	18,05	408
YRIMG1_1981	Personnes dont l'année d'immigration est entre 1981 et 1990	634	548	13,55	86
YRIMG1_1991	Personnes dont l'année d'immigration est entre 1991 et 2000	1 934	1 612	16,65	322
YRIMG1_2001	Personnes dont l'année d'immigration est entre 2001 et 2010	5 692	5 092	10,54	600

ADA = Aire de diffusion agrégée.
Source : Statistique Canada, échantillon du questionnaire détaillé du Recensement de 2016.

Ce tableau présente les contraintes ayant été utilisées dans le processus de calage pour Wood Buffalo et celles qui en ont été exclues. Une valeur de « Oui » signifie que la variable a été utilisée dans le processus, tandis qu'une valeur de « Non » indique que la variable en a été exclue.

Tableau D.2
Contraintes de calage exclues à Wood Buffalo

Variable/contrainte	Exclues à Wood Buffalo
ADULTCF	Non
AGE00_14	Oui
AGE14	Oui
AGE15_24	Oui
AGE19	Oui
AGE24	Oui
AGE25_34	Oui
AGE29	Oui
AGE34	Oui
AGE35_44	Oui
AGE39	Oui
AGE4	Oui
AGE44	Oui
AGE45_54	Oui
AGE49	Oui
AGE54	Oui
AGE55_64	Oui
AGE59	Oui
AGE64	Oui
AGE65PL	Oui
AGE74	Oui
AGE75PL	Oui
AGE9	Oui
APTLT5	Oui
CHILD	Non
CHILDFAM	Non
COMLAW_YE	Non
COMLAWNO_DIV	Non
COMLAWNO_OTHERS	Non
COMLAWNO_SEP	Non
COMLAWNO_SINGLE	Non
COMLAWNO_WID	Non
COMLAWYE_MARRIED	Non
COUPLE	Non
EMPIN_GT50	Oui
EMPIN_LE50	Oui
EMPIN_P0	Oui

Tableau D.2
Contraintes de calage exclues à Wood Buffalo

Variable/contrainte	Exclues à Wood Buffalo
EMPIN_P100	Oui
EMPIN_P25	Oui
EMPIN_P50	Oui
EMPIN_P75	Oui
FEMALE	Oui
FEMALEGE15	Oui
FEMALELT15	Oui
HHADACSD (ADA)	Oui
HHINC_GT50	Non
HHINC_LE50	Non
HHINC_P100	Non
HHINC_P25	Non
HHINC_P50	Non
HHINC_P75	Non
HHSIZE1	Non
HHSIZE2	Non
HHSIZE3	Non
HHSIZE4	Non
HHSIZE5	Non
HHSIZEGE5	Non
HHSIZEGE6	Non
INEFAM	Non
IR_LINK_NO	Non
IR_LINK_YE	Non
LIM_NO	Non
LIM_YE	Non
LONEPAR	Non
MALE	Oui
MALEGE15	Oui
MALELT15	Oui
MARRIED	Oui
NOCLDFAM	Non
NOINEFAM	Non
NOTINFAM	Non
OLN_BI	Oui
OLN_EN	Oui
OLN_FR	Oui
OLN_NO	Oui
POBG2_1	Non
POBG2_10	Non

Tableau D.2
Contraintes de calage exclues à Wood Buffalo

Variable/contrainte	Exclues à Wood Buffalo
POBG2_11	Non
POBG2_12	Non
POBG2_13	Non
POBG2_14	Non
POBG2_15	Non
POBG2_16	Non
POBG2_17	Non
POBG2_18	Non
POBG2_19	Non
POBG2_2	Non
POBG2_20	Non
POBG2_21	Non
POBG2_22	Non
POBG2_23	Non
POBG2_24	Non
POBG2_25	Non
POBG2_26	Non
POBG2_27	Non
POBG2_28	Non
POBG2_29	Non
POBG2_3	Non
POBG2_30	Non
POBG2_31	Non
POBG2_32	Non
POBG2_33	Non
POBG2_34	Non
POBG2_35	Non
POBG2_36	Non
POBG2_37	Non
POBG2_38	Non
POBG2_39	Non
POBG2_4	Non
POBG2_40	Non
POBG2_41	Non
POBG2_42	Non
POBG2_43	Non
POBG2_44	Non
POBG2_45	Non
POBG2_46	Non
POBG2_47	Non

Tableau D.2
Contraintes de calage exclues à Wood Buffalo

Variable/contrainte	Exclues à Wood Buffalo
POBG2_48	Non
POBG2_49	Non
POBG2_5	Non
POBG2_50	Non
POBG2_51	Non
POBG2_54	Non
POBG2_55	Non
POBG2_56	Non
POBG2_57	Non
POBG2_58	Non
POBG2_59	Non
POBG2_6	Non
POBG2_60	Non
POBG2_61	Non
POBG2_62	Non
POBG2_63	Non
POBG2_64	Non
POBG2_65	Non
POBG2_66	Non
POBG2_67	Non
POBG2_68	Non
POBG2_69	Non
POBG2_7	Non
POBG2_70	Non
POBG2_71	Non
POBG2_8	Non
POBG2_9	Non
POBG3_1	Non
POBG3_10	Non
POBG3_12	Non
POBG3_14	Non
POBG3_15	Non
POBG3_16	Non
POBG3_17	Non
POBG3_18	Non
POBG3_19	Non
POBG3_2	Non
POBG3_20	Non
POBG3_21	Non
POBG3_22	Non

Tableau D.2
Contraintes de calage exclues à Wood Buffalo

Variable/contrainte	Exclues à Wood Buffalo
POBG3_3	Non
POBG3_4	Non
POBG3_5	Non
POBG3_6	Non
POBG3_7	Non
POBG3_8	Non
POBG3_9	Non
PPADACSD (ADA)	Oui
SNGLDET	Oui
TOTCFAM	Non
TOTHHLD	Oui
TOTPERS	Oui
TPERGE15	Oui
TPERLT15	Oui
YRIMD_1900	Non
YRIMD_1981	Non
YRIMD_1986	Non
YRIMD_1991	Non
YRIMD_1996	Non
YRIMD_2001	Non
YRIMD_2006	Non
YRIMD_2011	Non
YRIMG1_1900	Non
YRIMG1_1981	Non
YRIMG1_1991	Non
YRIMG1_2001	Non
YRIMG1_2011	Non

Notes :
Oui indique que la variable a été utilisée au cours du processus.
Non indique que la variable a été exclue du processus.

Bibliographie

BUREAU FÉDÉRAL DE LA STATISTIQUE. 1968. *Sampling in the Census*, rapport interne, Statistique Canada, Ottawa, Ontario.

DEVIN, N., et F. VERRET. 2016. « The development of a variance estimation methodology for large-scale dissemination of quality indicators for the 2016 Canadian census long form sample », *JSM Proceedings*, Survey Research Methods Section, American Statistical Association, Chicago, États-Unis.

FELLEGI, I.P. 1964. « Response variance and its estimation », *Journal of the American Statistical Association*, vol. 59, nº 308, p. 1016 à 1041.

FOLSOM, R.E., Jr., et A.C. SINGH. 2000. « The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification », *Proceedings of the Survey Research Methods Section*, American Statistical Association, p. 598 à 603.

HANSEN, M.H., W.N. HURWITZ et M.A. BERSHAD. 1959. « Measurement errors in censuses and surveys », *Bulletin of the International Statistical Institute*, vol. 38, p. 359 à 374.

JUDKINS, D.R. 1990. « Fay's method for variance estimation », *Journal of Official Statistics*, Statistique Suède, vol. 6, nº 3, p. 223 à 239.

RAO, J.N.K., et J. Shao. 1999. « Modified balanced repeated replication for complex survey data », *Biometrika*, Oxford University Press, vol. 86, nº 2, p. 403 à 415.

ROYCE, D. 1983. *The Use of Sampling in the 1981 Canadian Census*, rapport interne, Statistique Canada, Ottawa, Ontario.

SHAO, J., et Q. TANG. 2011. « Random group variance estimators for survey data with random hot deck imputation », *Journal of Official Statistics*, Statistique Suède, vol. 27, nº 3, p. 507 à 526.

STATISTIQUE CANADA. 2014. Système canadien de vérification et d'imputation du recensement (SCANVIR), version 5.2, guide de base de l'utilisateur, Ottawa, Ontario.

STATISTIQUE CANADA. 2015. *Rapport technique sur l'échantillonnage et la pondération, Enquête nationale auprès des ménages (ENM) de 2011*, produit nº 99-002-X2011001 au catalogue, Ottawa, Ontario, version révisée le 24 mars 2015, http://www12.statcan.gc.ca/nhs-enm/2011/ref/reports-rapports/sw-ep/index-fra.cfm (site consulté le 28 février 2018).

STATISTIQUE CANADA. 2017. *Guide du Recensement de la population, 2016*, produit nº 98-304-X2016001 au catalogue, Ottawa, Ontario, version mise à jour le 30 janvier 2018, http://www12.statcan.gc.ca/census-recensement/2016/ref/98-304/index-fra.cfm (site consulté le 28 février 2018).